Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Complexity 22 (2006) 549–603
Journal of COMPLEXITY
www.elsevier.com/locate/jco

# Optimal approximation of elliptic problems by linear and nonlinear mappings II

Stephan Dahlke[a,1], Erich Novak[b,*], Winfried Sickel[b]

[a] *FB12 Mathematik und Informatik, Philipps-Universität Marburg, Hans-Meerwein Straße, Lahnberge, 35032 Marburg, Germany*
[b] *Mathematisches Institut, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany*

Received 13 September 2005; accepted 4 April 2006
Available online 14 June 2006

**Abstract**

We study the optimal approximation of the solution of an operator equation $\mathscr{A}(u) = f$ by four types of mappings: (a) linear mappings of rank $n$; (b) $n$-term approximation with respect to a Riesz basis; (c) approximation based on linear information about the right-hand side $f$; (d) continuous mappings. We consider worst case errors, where $f$ is an element of the unit ball of a Sobolev or Besov space $B^r_q(L_p(\Omega))$ and $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain; the error is always measured in the $H^s$-norm. The respective widths are the linear widths (or approximation numbers), the nonlinear widths, the Gelfand widths, and the manifold widths. As a technical tool, we also study the Bernstein numbers. Our main results are the following. If $p \geqslant 2$ then the order of convergence is the same for all four classes of approximations. In particular, the best linear approximations are of the same order as the best nonlinear ones. The best linear approximation can be quite difficult to realize as a numerical algorithm since the optimal Galerkin space usually depends on the operator and on the shape of the domain $\Omega$. For $p < 2$ there is a difference, nonlinear approximations are better than linear ones. However, in this case, it turns out that linear information about the right-hand side $f$ is again optimal. Our main theoretical tool is the best $n$-term approximation with respect to an optimal Riesz basis and related nonlinear widths. These general results are used to study the Poisson equation in a polygonal domain. It turns out that best $n$-term wavelet approximation is (almost) optimal. The main results

* Corresponding author.
*E-mail addresses:* dahlke@mathematik.uni-marburg.de (S. Dahlke), novak@math.uni-jena.de (E. Novak), sickel@math.uni-jena.de (W. Sickel)
*URLs:* http://www.mathematik.uni-marburg.de/~dahlke (S. Dahlke), http://www.minet.uni-jena.de/~novak (E. Novak), http://www.minet.uni-jena.de/~sickel (W. Sickel).

[1] The work of this author has been supported through the European Union's Human Potential Programme, under contract HPRN-CT-2002-00285 (HASSIP), and through DFG, Grant Da 360/4-2, Da 360/4-3.


doi:10.1016/j.jco.2006.04.001

of this paper are about approximation, not about computation. However, we also discuss consequences of the results for the numerical complexity of operator equations.


*MSC:* 41A25; 41A46; 41A65; 42C40; 65C99

*Keywords:* Elliptic operator equations; Worst case error; Linear and nonlinear approximation methods; Best *n*-term approximation; Besov spaces; Gelfand widths; Bernstein widths; Manifold widths

## 1. Introduction

We study the optimal approximation of the solution of an operator equation

$$\mathcal{A}(u) = f, \tag{1}$$

where $\mathcal{A}$ is a linear operator

$$\mathcal{A} : H \to G \tag{2}$$

from a Hilbert space $H$ to another Hilbert space $G$. We always assume that $\mathcal{A}$ is boundedly invertible, and so (1) has a unique solution for any $f \in G$. We have in mind the more specific situation of an elliptic operator equation which is given as follows. Assume that $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain and assume that

$$\mathcal{A} : H_0^s(\Omega) \to H^{-s}(\Omega) \tag{3}$$

is an isomorphism, where $s > 0$. (For the definition of the Sobolev spaces $H_0^s(\Omega)$ and $H^{-s}(\Omega)$, we refer to Sections A.7–A.9.) A standard case (for second-order elliptic boundary value problems for PDEs) is $s = 1$, but also other values of $s$ are of interest. Now we put $H = H_0^s(\Omega)$ and $G = H^{-s}(\Omega)$. Since $\mathcal{A}$ is boundedly invertible, the inverse mapping $S : G \to H$ is well defined. This mapping is sometimes called the solution operator—in particular if we want to compute the solution $u = S(f)$ from the given right-hand side $\mathcal{A}(u) = f$.

We use linear and (different kinds of) nonlinear mappings $S_n$ for the approximation of the solution $u = \mathcal{A}^{-1}(f)$ for $f$ contained in $F \subset G$. We consider the worst case error

$$e(S_n, F, H) = \sup_{\|f\|_F \leqslant 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H, \tag{4}$$

where $F$ is a normed (or quasi-normed) subspace of $G$. In our main results, $F$ will be a Sobolev or Besov space.[2] Hence we use the following commutative diagram

$$\begin{array}{ccc} G & \xrightarrow{S} & H \\ I \nwarrow & & \nearrow S_F \\ & F. & \end{array}$$

[2] Formally we only deal with Besov spaces. Because of the embeddings $B_1^{-s+t}(L_p(\Omega)) \subset W_p^{-s+t}(\Omega) \subset B_\infty^{-s+t}(L_p(\Omega))$, which hold for $1 \leqslant p \leqslant \infty$, $t \geqslant s$, see [91], our results are valid also for Sobolev spaces.

Here $I : F \to G$ denotes the identity and $S_F$ the restriction of $S$ to $F$. In the specific case (3) this diagram is given by

$$\begin{array}{ccc} H^{-s}(\Omega) & \xrightarrow{S} & H_0^s(\Omega) \\ I \quad \nwarrow & & \nearrow \quad S_t \\ & B_q^{-s+t}(L_p(\Omega)), & \end{array}$$

where $B_q^{-s+t}(L_p(\Omega))$ denotes a Besov space compactly embedded into $H^{-s}(\Omega)$, cf. the Appendix for a definition, and $S_t$ the restriction of $S$ to $B_q^{-s+t}(L_p(\Omega))$. We are interested in approximations that have the optimal order of convergence depending on $n$, where $n$ denotes the *degree of freedom*. In general our results are *constructive in a mathematical sense*, because we can describe optimal approximations $S_n$ in mathematical terms. This does not mean, however, that these descriptions are constructive in a practical sense, since it might be very difficult to convert those descriptions into a practical algorithm. We will discuss this more thoroughly in Section 3.4. As a consequence, most of our results give optimal benchmarks and can serve for the evaluation of old and new algorithms. We study and compare *four kinds of approximation methods*; see Section 2.1 for details.

- We consider the class $\mathcal{L}_n$ of all continuous linear mappings $S_n : F \to H$,

  $$S_n(f) = \sum_{i=1}^{n} L_i(f) \cdot \tilde{h}_i$$

  with arbitrary $\tilde{h}_i \in H$. The worst case error of optimal linear mappings is given by the *approximation numbers* or *linear widths*

  $$e_n^{\mathrm{lin}}(S, F, H) = \inf_{S_n \in \mathcal{L}_n} e(S_n, F, H).$$

- For a given basis $\mathcal{B}$ of $H$ we consider the class $\mathcal{N}_n(\mathcal{B})$ of all (linear or nonlinear) mappings of the form

  $$S_n(f) = \sum_{k=1}^{n} c_k \, h_{i_k},$$

  where the $c_k$ and the $i_k$ depend in an arbitrary way on $f$. We also allow that the basis $\mathcal{B}$ to be chosen in a nearly arbitrary way. Then the *nonlinear widths* $e_{n,C}^{\mathrm{non}}(S, F, H)$ are given by

  $$e_{n,C}^{\mathrm{non}}(S, F, H) = \inf_{\mathcal{B} \in \mathcal{B}_C} \inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, F, H).$$

  Here $\mathcal{B}_C$ denotes a set of Riesz bases for $H$ where $C$ indicates the stability of the basis. These numbers are the main topic of our analysis.
- We also study methods $S_n$ with $S_n = \varphi_n \circ N_n$, where $N_n : F \to \mathbb{R}^n$ is linear and continuous and $\varphi_n : \mathbb{R}^n \to H$ is arbitrary. This is the class of all (linear or nonlinear) approximations $S_n$ that use *linear information of cardinality n* about the right-hand side $f$. The respective widths are

  $$r_n(S, F, H) := \inf_{S_n} e(S_n, F, H),$$

  they are closely related to the *Gelfand numbers*.

- Let $\mathcal{C}_n$ be the class of continuous mappings, given by arbitrary continuous mappings $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$. Again we define the worst case error of optimal continuous mappings by

$$e_n^{\text{cont}}(S, F, H) = \inf_{S_n \in \mathcal{C}_n} e(S_n, F, H),$$

where $S_n = \varphi_n \circ N_n$. These numbers are called *manifold widths* of $S$.

For problems (3) with $F = B_q^r(L_p(\Omega))$ our main results are the following. If $p \geqslant 2$ then the order of convergence is the same for all four classes of approximations. In particular, the best linear approximations are of the same order as the best nonlinear ones. The best linear approximation can be quite difficult to realize as a numerical algorithm since the optimal Galerkin space usually depends on the operator and on the shape of the domain $\Omega$. For $p < 2$ there is an essential difference, nonlinear approximations are better than linear ones. However, in this case it turns out that linear information about the right-hand side $f$ is optimal. Our main theoretical tool is best $n$-term approximation with respect to an optimal Riesz basis and related nonlinear widths. The main results are about approximation, not about computation. However, we also discuss consequences of the results for the numerical complexity of operator equations.

The paper is organized as follows:

1. Introduction.
2. Linear and nonlinear widths.
2.1. Classes of admissible mappings.
2.2. Properties of widths and relations between them.
3. Optimal approximation of elliptic problems.
3.1. Optimal linear approximation of elliptic problems.
3.2. Optimal nonlinear approximation of elliptic problems.
3.3. The Poisson equation.
3.4. Algorithms and complexity.
4. Proofs.
4.1. Properties of widths.
4.2. Widths of embeddings of weighted sequence spaces.
4.3. Widths of embeddings of Besov spaces.
4.4. Proofs of Theorems 2, 3, and 5.
5. Appendix—Besov spaces.

We add a few comments. The main results of our paper are contained in Section 3.2. They are further illustrated for the case of the Poisson equation in Section 3.3. A discussion in connection with *uniform approximation*, *adaptive/nonadaptive information*, *adaptive numerical schemes*, and *complexity* is contained in Section 3.4. All proofs are contained in Section 4. Of independent interest are the estimates of the widths of embedding operators for Besov spaces, see Section 4.3.

**Notation.** We write $a \asymp b$ if there exists a constant $c > 0$ (independent of the context-dependent relevant parameters) such that

$$c^{-1}\, a \leqslant b \leqslant c\, a\,.$$

All unimportant constants will be denoted by $c$, sometimes with additional indices.

## 2. Linear and nonlinear widths

Widths represent concepts of optimality. In this section we shall discuss several variants. Most important for us will be the nonlinear widths $e_n^{\text{non}}$ and the linear widths $e_n^{\text{lin}}$. We also study Gelfand and manifold widths and, as a vehicle of the proofs, Bernstein widths.

### 2.1. *Classes of admissible mappings*

#### 2.1.1. *Linear mappings $S_n$*

Here we consider the class $\mathcal{L}_n$ of all continuous linear mappings $S_n : F \to H$,

$$S_n(f) = \sum_{i=1}^{n} L_i(f) h_i \tag{5}$$

where the $L_i : F \to \mathbb{R}$ are linear functionals and $h_i$ are elements of $H$. We consider the worst case error

$$e(S_n, F, H) := \sup_{\|f\|_F \leqslant 1} \|\mathcal{A}^{-1}(f) - S_n(f)\|_H, \tag{6}$$

where $F$ is a normed (or quasi-normed) subspace of $G$. Accordingly, we seek the optimal linear approximation, as well as the numbers

$$e_n^{\text{lin}}(S, F, H) = \inf_{S_n \in \mathcal{L}_n} e(S_n, F, H), \tag{7}$$

usually called *approximation numbers* or *linear widths* of $S : F \to H$, cf. [60,72,73,85].

#### 2.1.2. *Nonlinear mappings $S_n$*

Let $\mathcal{B} = \{h_1, h_2, \ldots\}$ be a subset of $H$. Then the *best n-term approximation* of an element $u \in H$ with respect to this set $\mathcal{B}$ is defined as

$$\sigma_n(u, \mathcal{B})_H := \inf_{i_1, \ldots, i_n} \inf_{c_1, \ldots c_n} \left\| u - \sum_{k=1}^{n} c_k \, h_{i_k} \right\|_H . \tag{8}$$

This subject is widely studied, see the surveys [29,84]. Now we continue by looking for an optimal set $\mathcal{B}$ as has been done in [33,38,54,82–84]. Temlyakov [84] suggested to consider the quantities

$$\inf_{\mathcal{B} \in \mathcal{D}} \sup_{\|u\|_Y \leqslant 1} \sigma_n(u, \mathcal{B})_H,$$

where $\mathcal{D}$ is a subset of the set of all bases of $H$. The particular case of $\mathcal{D}$ being the set of all orthonormal bases has been discussed in [82,83], while the set of all unconditional, democratic bases is studied in [33]. See Remark 25 for a further discussion. In this paper we work with Riesz bases, see, e.g., [62, p. 21].

**Definition 1.** Let $H$ be a Hilbert space. Then the sequence $h_1, h_2, \ldots$ of elements of $H$ is called a *Riesz basis* for $H$ if there exist positive constants $A$ and $B$ such that, for every sequence of scalars

$\alpha_1, \alpha_2, \ldots$ with $\alpha_k \neq 0$ for only finitely many $k$, we have

$$A \left( \sum_k |\alpha_k|^2 \right)^{1/2} \leqslant \left\| \sum_k \alpha_k h_k \right\|_H \leqslant B \left( \sum_k |\alpha_k|^2 \right)^{1/2} \tag{9}$$

and the vector space of finite sums $\sum \alpha_k h_k$ is dense in $H$.

**Remark 1.** The constants $A, B$ reflect the stability of the basis. Orthonormal bases are those with $A = B = 1$. Typical examples of Riesz bases are the biorthogonal wavelet bases on $\mathbb{R}^d$ or on certain Lipschitz domains, cf. [12, Sections 2.6, 2.12].

In what follows

$$\mathcal{B} = \{h_i \mid i \in \mathbb{N}\} \tag{10}$$

will always denote a Riesz basis of $H$ with $A$ and $B$ being the corresponding optimal constants in (9).

For a given basis $\mathcal{B}$ we consider the class $\mathcal{N}_n(\mathcal{B})$ of all (linear or nonlinear) mappings of the form

$$S_n(f) = \sum_{k=1}^{n} c_k h_{i_k}, \tag{11}$$

where the $c_k$ and the $i_k$ depend in an arbitrary way on $f$. By the arbitrariness of $S_n$ one obtains immediately

$$\inf_{S_n \in \mathcal{N}_n(\mathcal{B})} \sup_{\|f\|_F \leqslant 1} \|\mathcal{A}^{-1} f - S_n(f)\|_H = \sup_{\|f\|_F \leqslant 1} \sigma_n(\mathcal{A}^{-1} f, \mathcal{B})_H. \tag{12}$$

It is natural to assume some common stability of the bases under consideration. For a real number $C \geqslant 1$ we put

$$\mathcal{B}_C := \Big\{ \mathcal{B} : B/A \leqslant C \Big\}. \tag{13}$$

We are ready to define the nonlinear widths $e_{n,C}^{\mathrm{non}}(S, F, H)$ by

$$e_{n,C}^{\mathrm{non}}(S, F, H) = \inf_{\mathcal{B} \in \mathcal{B}_C} \inf_{S_n \in \mathcal{N}_n(\mathcal{B})} e(S_n, F, H). \tag{14}$$

These numbers are the main topic of our analysis. We call them the *widths of best n-term approximation* (with respect to the collection $\mathcal{B}_C$ of Riesz bases of $H$).

**Remark 2.** (i) It should be clear that the class $\mathcal{N}_n(\mathcal{B})$ contains many mappings that are difficult to compute. In particular, the number $n$ just reflects the dimension of a nonlinear manifold and has nothing to do with a computational cost. In this paper we also are interested in lower bounds, such lower bounds being strengthened if we admit a larger class of approximations.

(ii) The inequality

$$e_{n,C}^{\mathrm{non}}(S, F, H) \leqslant e_n^{\mathrm{lin}}(S, F, H) \tag{15}$$

is trivial.

(iii) Because of the homogeneity of $\sigma_n$, i.e., $\sigma_n(\lambda u, \mathcal{B})_H = |\lambda| \, \sigma_n(u, \mathcal{B})_H$, $\lambda \in \mathbb{R}$, it does not change the asymptotic behavior of $e_n^{\mathrm{non}}$ if we replace $\sup_{\|f\|_F \leqslant 1}$ by $\sup_{\|f\|_F \leqslant c}$ for $c > 0$.

*2.1.3. Continuous mappings $S_n$*

Linear mappings $S_n$ are of the form $S_n = \varphi_n \circ N_n$ where both $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$ are linear and continuous. If we drop the linearity condition then we obtain the class of all continuous mappings $\mathcal{C}_n$, given by arbitrary continuous mappings $N_n : F \to \mathbb{R}^n$ and $\varphi_n : \mathbb{R}^n \to H$. Again we define the worst case error of optimal continuous mappings by

$$e_n^{\mathrm{cont}}(S, F, H) = \inf_{S_n \in \mathcal{C}_n} e(S_n, F, H). \tag{16}$$

These numbers, or slightly different numbers, were studied by different authors, cf. [30,32,40,60]. Sometimes these numbers are called *manifold widths* of $S$, see [32], and we will use this terminology here. The inequality

$$e_n^{\mathrm{cont}}(S, F, H) \leqslant e_n^{\mathrm{lin}}(S, F, H) \tag{17}$$

is obvious.

*2.1.4. Gelfand widths and minimal radii of information*

We can also study methods $S_n$ with $S_n = \varphi_n \circ N_n$, where $N_n : F \to \mathbb{R}^n$ is linear and continuous and $\varphi_n : \mathbb{R}^n \to H$ is arbitrary. The respective widths are

$$r_n(S, F, H) := \inf_{S_n} e(S_n, F, H). \tag{18}$$

These numbers are called the $n$th *minimal radii of information*, which are closely related to Gelfand widths, see Lemma 1. The *nth Gelfand width* of the linear operator $S : F \to H$ is given by

$$d^n(S, F, H) := \inf_{L_1,\dots,L_n} \sup\{\|Sf\|_H \ : \ \|f\|_F \leqslant 1,\ L_i(f) = 0\,,\ i = 1, \dots n\}\,, \tag{19}$$

where the $L_i : F \to \mathbb{R}$ are continuous linear functionals.

*2.1.5. Bernstein widths*

A well-known tool for deriving lower bounds of widths consists in the investigation of Bernstein widths, see [72,73,85].

**Definition 2.** The number $b_n(S, F, H)$, called the $n$th *Bernstein width* of the operator $S : F \to H$, is the radius of the largest $(n + 1)$-dimensional ball that is contained in $S(\{\|f\|_F \leqslant 1\})$.

**Remark 3.** The literature contains several different definitions of Bernstein widths. For example, Pietsch [71] gives the following version. Let $X_n$ denote subspaces of $F$ of dimension $n$. Then

$$\widetilde{b}_n(S, F, H) := \sup_{X_n \subset F} \inf_{x \in X_n, x \neq 0} \frac{\|Sx\|_H}{\|x\|_F}\,.$$

As long as $S$ is an injective mapping we obviously have $b_n(S, F, H) = \widetilde{b}_{n+1}(S, F, H)$.

*2.2. Properties of widths and relations between them*

**Lemma 1.** *Let $n \in \mathbb{N}$ and assume that $F \subset G$ is quasi-normed.*

(i) *We have $d^n \leqslant r_n \leqslant 2d^n$ if $F$ is normed and $d^n \asymp r_n$ in general.*

(ii) *The inequality*

$$b_n(S, F, H) \leqslant \min\left(e_n^{\text{cont}}(S, F, H), d^n(S, F, H)\right) \tag{20}$$

*holds for all* $n$.

**Remark 4.** The inequality $b_n \leqslant e_n^{\text{cont}}$ is known, compare e.g., with [30], and the proof technique (via Borsuk's theorem) is often used for the proof of similar results.

The Bernstein widths $b_n$ can also be used to prove lower bounds for the $e_{n,C}^{\text{non}}$. The following inequality has been proved in [24].

**Lemma 2.** *Assume that* $F \subset G$ *is quasi-normed. Then*

$$e_{n,C}^{\text{non}}(S, F, H) \geqslant \frac{1}{2C} b_m(S, F, H) \tag{21}$$

*holds for all* $m \geqslant 4C^2 n$.

More important for us will be a direct comparison of $e_n^{\text{non}}$ and $e_n^{\text{cont}}$. Best $n$-term approximation yields a mapping

$$S_n(u) = \sum_{k=1}^{n} c_k h_{i_k}$$

which is in general not continuous. However, it is known that certain discontinuous mappings can be suitably modified in order to obtain a continuous $n$-term approximation with an error which is only slightly worse, see, for example, [32,41]. We prove that, under general assumptions, the numbers $e_{n,C}^{\text{non}}$ can be bounded from below by the manifold widths $e_n^{\text{cont}}$.

**Theorem 1.** *Let* $S : G \to H$ *be an isomorphism. Suppose that the embedding* $F \hookrightarrow G$ *is compact. Then for all* $C \geqslant 1$ *and all* $n \in \mathbb{N}$, *we have*

$$e_{4n+1}^{\text{cont}}(S, F, H) \leqslant 2C \|S\|^2 \|S^{-1}\|^2 \, e_{n,C}^{\text{non}}(S, F, H). \tag{22}$$

Finally, we collect some further properties of the quantities $e_n^{\text{cont}}$ and $e_n^{\text{non}}$.

**Lemma 3.** (i) *Let* $m, n \in \mathbb{N}$, *and let* $F$ *be a subset of the quasi-normed linear space* $X$, *where* $X$ *itself is a subset of the quasi-normed linear space* $Y$. *Let* $I_j$ *denote embedding operators. Then*

$$e_{m+n}^{\text{cont}}(I_1, F, Y) \leqslant e_m^{\text{cont}}(I_2, F, X)\, e_n^{\text{cont}}(I_3, X, Y) \tag{23}$$

*holds.*

(ii) *Let* $F$ *be a quasi-normed subset of* $G$ *and let* $I : F \to G$ *be the embedding. Then*

$$e_n^{\text{cont}}(I, F, G) \leqslant \|S^{-1}\|\, e_n^{\text{cont}}(S, F, H) \leqslant \|S^{-1}\| \, \|S\|\, e_n^{\text{cont}}(I, F, G) \tag{24}$$

*and for any* $C \geqslant \|S^{-1}\| \, \|S\|$, *we have*

$$\begin{aligned} e_{n,C\,\|S^{-1}\|\,\|S\|}^{\text{non}}(I, F, G) &\leqslant \|S^{-1}\|\, e_{n,C}^{\text{non}}(S, F, H) \\ &\leqslant \|S^{-1}\| \, \|S\|\, e_{n,C/(\|S^{-1}\|\,\|S\|)}^{\text{non}}(I, F, G). \end{aligned} \tag{25}$$

**Remark 5.** Let us point out the following which is part of the proof of Lemma 3. Let $\mathcal{B} = \{h_1, h_2, \dots\}$ be a Riesz basis of $G$. Let $S_n$ be an approximation of the identity $I : F \to G$. Then $S(\mathcal{B})$ is a Riesz basis of $H$ and $S \circ S_n$ is an approximation of $S : F \to H$ satisfying

$$\|f - S_n(f)\|_G \leqslant \|S^{-1}\| \cdot \|Sf - S \circ S_n(f)\|_H \leqslant \|S^{-1}\| \cdot \|S\| \cdot \|f - S_n(f)\|_G. \tag{26}$$

This makes clear that if $\mathcal{B}$ and $S_n$ are order optimal for the triple $I, F, G$, then $S(\mathcal{B})$ and $S \circ S_n$ are order optimal for the triple $S, F, H$. Consequently, instead of looking for good approximations of $S : F \to H$ it will be enough to study approximations of the embedding $I : F \to G$.

**Remark 6.** The assertion in part (i) of the lemma is essentially proved in [40] but traced there to Khodulev. The inequality (23) can be made more transparent by means of the diagram

$$\begin{array}{ccc} X & \xrightarrow{I_3} & Y \\ I_2 \nwarrow & & \nearrow I_1 \\ & F. & \end{array}$$

**Remark 7.** The approximation numbers $e_n^{\mathrm{lin}}$, the Gelfand widths $d^n$, the manifold widths $e_n^{\mathrm{cont}}$ and the Bernstein widths $b_n$ are particular examples of $s$-numbers in the sense of Pietsch [71], see [60] for the manifold widths. They have several properties in common. Letting $s_n$ denote any of the numbers $e_n^{\mathrm{lin}}$, $d^n$, $e_n^{\mathrm{cont}}$ and $b_n$, we have

$$s_n(T_2 \circ T_1 \circ T_0) \leqslant \|T_0\| \|T_2\| \, s_n(T_1), \tag{27}$$

where $T_0 \in \mathcal{L}(E_0, E)$, $T_1 \in \mathcal{L}(E, F)$, $T_2 \in \mathcal{L}(F, F_0)$ and $E_0, E, F, F_0$ are arbitrary Banach spaces. For these four types of $s$-numbers the assertion remains true also for quasi-Banach spaces.

Another property concerns additivity. For $s_n$ instead of $e_n^{\mathrm{lin}}$ and $d^n$ we have

$$s_{2n}(T_0 + T_1) \leqslant c \left( s_n(T_0) + s_n(T_1) \right), \tag{28}$$

where $T_0, T_1 \in \mathcal{L}(E, F)$, $E, F$ are arbitrary quasi-Banach spaces, and $c$ does not depend on $n, T_0, T_1$, cf. [10]. In case that $F$ is a Banach space, one can take $c = 1$.

## 3. Optimal approximation of elliptic problems

Let $s, t > 0$. We consider the diagram

$$\begin{array}{ccc} H^{-s}(\Omega) & \xrightarrow{S} & H_0^s(\Omega) \\ I \nwarrow & & \nearrow S_t \\ & B_q^{-s+t}(L_p(\Omega)), & \end{array}$$

where $S_t$ denotes the restriction of $S$ to $B_q^{-s+t}(L_p(\Omega))$ and $I$ denotes the identity. We assume (3) and we let $S = \mathcal{A}^{-1}$.

### 3.1. Optimal linear approximation of elliptic problems

**Theorem 2.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $0 < p, q \leqslant \infty$, $s > 0$, and*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+. \tag{29}$$

*Then*

$$e_n^{\mathrm{lin}}(S, B_q^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \asymp \begin{cases} n^{-t/d} & \text{if } 2 \leqslant p \leqslant \infty, \\ n^{-t/d+1/p-1/2} & \text{if } 0 < p < 2. \end{cases}$$

**Remark 8.** (i) The restriction (29) is necessary and sufficient for the compactness of the embedding $I : B_q^{-s+t}(L_p(\Omega)) \hookrightarrow H^{-s}(\Omega)$, cf. the Appendix, Proposition 7.

(ii) The proof is constructive. First of all one has to determine a linear mapping $S_n$ that approximates the embedding $I : B_q^{-s+t}(L_p(\Omega)) \to H^{-s}(\Omega)$ with the optimal order. How this can be done is described in Remark 28, Section 4.3.3. Finally, the linear mapping $S \circ S_n$ realizes an order-optimal approximation of $S_t$.

(iii) There are hundreds of references dealing with approximation numbers of linear operators. Most useful for us have been the monographs [43,61,72,73,81,85,94], as well as the references contained therein.

### 3.2. Optimal nonlinear approximation of elliptic problems

To begin with, we consider the manifold and the Gelfand widths. There we have a rather final answer.

**Theorem 3.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $0 < p, q \leqslant \infty$, $s > 0$, and*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+ .$$

*Then*

$$e_n^{\mathrm{cont}}(S, B_q^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \asymp n^{-t/d} .$$

*If, in addition, $p \geqslant 1$ (and $t > d/2$ if $1 \leqslant p < 2$), then*

$$d^n(S, B_q^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \asymp n^{-t/d}.$$

From Theorems 1 and 3 we conclude that the order of $e_{n,C}^{\mathrm{non}}$ is also at least $n^{-t/d}$. For the respective upper bound of the nonlinear widths $e_{n,C}^{\mathrm{non}}$ we need a few more restrictions with respect to the domain $\Omega$. Let $\Omega$ be a bounded Lipschitz domain in $\mathbb{R}^d$ and let $s > 0$. We assume that for any fixed triple $(t, p, q)$ of parameters the spaces $B_q^{-s+t}(L_p(\Omega))$ and $H^{-s}(\Omega)$ allow a discretization by one common wavelet system $\mathcal{B}^*$, i.e., (107)–(112) should be satisfied with $B_q^{-s+t}(L_p(\Omega))$ and $B_2^{-s}(L_2(\Omega))$, respectively, cf. Appendix A.10. By assumption such a wavelet system belongs to $\mathcal{B}_{C^*}$ for some $1 \leqslant C^* < \infty$.

**Theorem 4.** *Under the above conditions on $\Omega$ and if $0 < p, q \leqslant \infty$, $s > 0$, $t > d(\frac{1}{p} - \frac{1}{2})_+$, we have for any $C \geqslant C^*$*

$$e_{n,C}^{\mathrm{non}}(S, B_q^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \asymp n^{-t/d}.$$

**Remark 9.** Comparing Theorems 2–4 there is a clear message. For $p < 2$ there are nonlinear approximations that are better in order than any linear approximation.

**Remark 10.** The proof of the upper bound in Theorem 4 is *constructive in a theoretical sense* that we now describe. Given a right-hand side $f \in B_q^{-s+t}(L_p(\Omega))$ we have to calculate all wavelet coefficients $\langle f, \widetilde{\psi}_{j,\lambda}\rangle$. The sequence of these coefficients belongs to the space $b_{p,q}^{-s+t}(\nabla)$, cf. Section 4.2. With

$$a = (a_{j,\lambda})_{j,\lambda}, \quad a_{j,\lambda} := \langle f, \widetilde{\psi}_{j,\lambda}\rangle \quad \text{for all } j, \lambda,$$

we find a good approximation $S_n(a)$ of $a$ with $n$ components with respect to the norm $\| \cdot | b_{2,2}^s(\nabla)\|$ in Proposition 2. To get an optimal approximation of the solution $u = Sf$ in $\| \cdot | H^s(\Omega)\|$ we have to apply the solution operator to $S_n(a)$. Hence

$$u_n = (S \circ S_n)(a) = \sum_{j=0}^{K} \sum_{\lambda \in \Lambda_j^*} a_{j,\lambda}^* \, S\psi_{j,\lambda}, \tag{30}$$

where $K = K(a, n)$, with $a_{j,\lambda}^*$ and $\Lambda_j^*$ as in Proposition 2 (cf. in particular (62) and (65)), represents such a good approximation of $u$. To calculate $u_n$, a lot of computations have to be done. The coefficients $a_{j,\lambda}^*$ are the largest in a weighted sense (the weight depends on $n$ and $j$, cf. the proof of Proposition 2 for explicit formulas). Having these coefficients at hand one has finally to solve all the equations

$$\mathcal{A}u_{j,\lambda} = \psi_{j,\lambda}, \quad 0 \leqslant j \leqslant K, \quad \lambda \in \Lambda_j^* \tag{31}$$

to obtain $u_{j,\lambda} = S\psi_{j,\lambda}$. The number of equations is $O(n)$.

In this way we obtain a nonlinear approximation with respect to the Riesz basis given by the $S\psi_{j,\lambda}$. Observe that this Riesz basis depends on the operator equation. It would be much better to use a known Riesz basis, such as a wavelet basis, that does not depend on $\mathcal{A}$. See Theorem 5 for a step in that direction.

**Remark 11.** At least if $\Omega$ is a cube, all required properties are known to be satisfied if in addition $1 < p, q < \infty$. The latter restriction allows to use duality arguments, cf. Proposition 10 in Appendix A.8. There also exist results for domains with piecewise analytic boundary such as polygonal or polyhedral domains. One natural way as, e.g., outlined in [8,26] is to decompose the domain into a disjoint union of parametric images of reference cubes. Then, one constructs wavelet bases on the reference cubes and glues everything together in a judicious fashion. However, due to the glueing procedure, only Sobolev spaces $H^s$ with smoothness $s < \frac{3}{2}$ can be characterized. This bottleneck can be circumvented by the approach in [27]. There, a much more tricky domain decomposition method involving certain projection and extension operators is used. By proceeding in this way, norm equivalences for all spaces $B_q^t(L_p(\Omega))$ can be derived, at least for the case $p > 1$, see [27, Theorem 3.4.3]. However, the authors also mention that their results can be generalized to the case $p < 1$, see [27, Remark 3.1.2].

Sobolev and Besov spaces on compact $C^\infty$-manifolds were already characterized via spline bases and sequence spaces by Ciesielski and Figiel [11]. In that paper also the isomorphism between function spaces and sequence spaces is used to obtain results for various $s$-numbers.

**Remark 12.** Comparing Theorems 3 and 4 we see that the numbers $e_{n,C}^{\text{non}}$, $e_n^{\text{cont}}$, and $d^n$ have the same asymptotic behavior, at least for $p > 1$. Using the relation $d^n \asymp r_n$, see Lemma 1, we actually can get the optimal order $n^{-t/d}$ with an approximation of the form

$$f \mapsto S \circ \varphi_n \circ N_n(f), \tag{32}$$

where

$$N_n : B_q^{-s+t}(L_p(\Omega)) \to \mathbb{R}^n$$

is linear (this mapping gives the *information* that is used about the right-hand side), and

$$\varphi_n : \mathbb{R}^n \to H^{-s}(\Omega)$$

is nonlinear. Note that neither $N_n$ nor $\varphi_n$ depend on $S$. The mapping $\varphi_n \circ N_n$ gives a good approximation of the embedding from $B_q^{-s+t}(L_p(\Omega))$ to $H^{-s}$.

**Remark 13.** There is a further little difference between linear and nonlinear approximation. Let us consider the limiting case $t = d(1/p - 1/2)$, where $0 < p < 2$. Then the embedding $B_p^{-s+t}(L_p(\Omega)) \hookrightarrow H^{-s}(\Omega)$ is continuous, not compact. As a consequence

$$e_n^{\mathrm{lin}}(S, B_p^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \not\to 0 \quad \text{if } n \to \infty,$$

but

$$e_n^{\mathrm{non}}(S, B_p^{-s+t}(L_p(\Omega)), H_0^s(\Omega)) \to 0 \quad \text{if } n \to \infty,$$

cf. Remark 26.

### *3.3. The Poisson equation*

The next step is to discuss the specific case of the Poisson equation on a Lipschitz domain $\Omega$ contained in $\mathbb{R}^2$:

$$\begin{aligned} -\triangle u &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{33}$$

As usual, we study (33) in the weak formulation. Then, it can be shown that the operator $\mathcal{A} = \triangle : H_0^1 \longrightarrow H^{-1}$ is boundedly invertible, see, e.g., [50] for details. Hence Theorems 2 and 3 apply with $s = 1$; for the upper bound of Theorem 4 we need some restrictions with respect to $\Omega$. For the proof of Theorem 4 we used the Riesz basis $S\psi_{j,\lambda}$, which depends on $\mathcal{A}$. Now we want to approximate the solution $u$ by wavelets.

We shall restrict ourselves to the case that $\Omega$ is a simply connected polygonal domain. The segments of $\partial\Omega$ are denoted by $\overline{\Gamma}_1, \dots, \overline{\Gamma}_N$, where each $\Gamma_l$ is open and the segments are numbered in positive orientation. Furthermore, $\Upsilon_l$ denotes the endpoint of $\Gamma_l$ and $\omega_l$ denotes the measure of the interior angle at $\Upsilon_l$. Appropriate wavelet systems can be constructed for such a domain, see Remark 11. Then we obtain the following.

**Theorem 5.** *Let $\Omega$ be a polygonal domain in $\mathbb{R}^2$. Let $1 < p \leqslant 2$ and let $k \geqslant 1$ be a non-negative integer such that*

$$\frac{m\pi}{\omega_l} \neq k + 1 - \frac{2}{p} \qquad \text{for all } m \in \mathbb{N},\ l = 1, \dots, N.$$

*Then for an appropriate wavelet system $\mathcal{B}^*$, the best n-term approximation of problem* (33) *yields*

$$\sup_{\|f|B_p^{k-1}(L_p(\Omega))\| \leqslant 1} \sigma_n(u, \mathcal{B}^*) \leqslant c_\varepsilon n^{-k/2+\varepsilon} \tag{34}$$

*where $\varepsilon > 0$ and $c_\varepsilon$ do not depend on n.*

**Remark 14.** This approximation differs greatly from the one described in Remark 10. Here we can work with one given wavelet system to approximate the solution $u$. We are not forced to work with the solutions of the system (31). A more detailed discussion of these relationships, including possible numerical realizations of wavelet methods, will follow in Section 3.4.

### *3.4. Algorithms and complexity*

So far, we have studied the error $e(S_n, F, H)$ of approximations $S_n$. We compared the error of nonlinear $S_n$ and linear $S_n$ and stated results on the optimal rate of convergence. We assume that (1) is a given fixed operator equation and hence, in the case of (3), also $\Omega$ is fixed.

In this section we briefly discuss algorithms and their complexity, and for simplicity we still assume that the operator equation (3) is given and fixed. Observe that in practice it is important to construct also algorithms for more general problems: we want to input information about $\Omega$ and $\mathcal{A}$ and the right-hand side $f$, and we want to obtain an $\varepsilon$-approximation of the solution $u$. In our more restricted case we *only* have to input information concerning the right-hand side $f$ because $\Omega$ and $\mathcal{A}$ are fixed.

As is usual in numerical analysis, we use the real number model of computation (see [64] for the details and [66,67] for further comments). Any algorithm computes and/or uses some information (consisting in finitely many numbers) describing the right-hand side $f$ of (3). There are different ways how an algorithm may use information concerning $f$, we describe two of them in turn:

1. The information used about $f$ is very explicit if $S_n$ is linear (5): then the algorithm uses $L_1(f), \dots, L_n(f)$ and we assume that we have an oracle (or subroutine) for the $L_i(f)$. In practical applications the computation of a functional $L_i(f)$ can be very easy or very difficult or anything between. One often assumes that the cost of obtaining a value $L_i(f)$ is $c$ where $c > 0$ is small or large, depending on the circumstances.

   As in (11), we can imagine $S_n$ as the input–output mapping of a numerical algorithm: on input $f \in F$ we obtain the output $S_n(f) = u_n = \sum_{k=1}^{n} c_k h_{i_k}$. More formally we should say that the output is

   $$\mathrm{out}(f) = (i_1, c_1, i_2, c_2, \dots, i_n, c_n) \tag{35}$$

   but we identify $\mathrm{out}(f)$ with $u_n$. Of course we cannot consider arbitrary mappings $S_n$ of the form (11) as the input–output mapping of an algorithm, since not all such $S_n$ are computable.

   We still assume that we only have an oracle for the computation of linear functionals $L_i(f)$. Then it is not so clear what the information cost of (11) is, since (11) only describes the (desired) output of an algorithm, it is not an algorithm by itself. We need an algorithm that uses information $L_1(f), \dots, L_N(f)$, where $N$ might be bigger than $n$, to produce the $i_k$ and the $c_k$ of $\mathrm{out}(f)$. The information cost of such a procedure would be $cN$.
2. One also can assume that a good approximation $f_n$ can easily be precomputed with negligible cost. Hence the algorithm starts with an approximation

   $$f_n = \sum_{k=1}^{n} c_k g_{i_k} \tag{36}$$

   such as a best $n$-term approximation (or a greedy approximation) of $f$ with respect to a basis $\{g_i, : i \in \mathbb{N}\}$.

This is a good place for a short remark about adaption. The use of *adaptive methods* is quite widespread but we want to stress that the notion of adaptive methods is not uniformly used in the literature. Some confusion is almost unavoidable if such different notions are mixed. To avoid such confusion, we do not use the notion of an "adaptive method". Instead we speak first about *adaptive* (*or nonadaptive*) *information* and then about *adaptive numerical schemes*.

- *Nonadaptive information*: The algorithm uses certain functionals $L_1, L_2, \dots, L_n$ and for each input $f \in F$ the algorithm needs $L_1(f), L_2(f), \dots, L_n(f)$. Hence the functionals $L_i$ do not depend on *f*. In this case we say that the algorithm uses *nonadaptive information*.
- *Adaptive information*: The algorithm uses $L_1(f)$ and, depending on this number, the next functional $L_2$ is chosen. In general, the chosen functional $L_k$ may depend on the values $L_1(f), \dots, L_{k-1}(f)$ that are already known to the algorithm. Observe that $L_k$ cannot depend in an arbitrary way on *f* since the algorithm can only use the known information about *f*. In this case we say that the algorithm uses *adaptive information*.

We give an example. Assume that a certain $S_n$ of the form (11) can be realized in such a way that we first compute $L_1(f), \dots, L_N(f)$, where the $L_i$ do not depend on $f \in F$. In the latter parts of the algorithm we only use the $L_i(f)$ for the *n* largest values of $|L_i(f)|$, together with the corresponding values of *i*, to compute the output $\mathrm{out}(f)$. Such an algorithm uses nonadaptive information (of cardinality *N*), the information cost is $cN$.

There is a large stream of results, giving conditions under which adaptive information is superior (or not superior) compared to nonadaptive information; we mention the pioneering paper by Bakhvalov [2], the results on operator equations by Gal and Micchelli [44] and by Traub and Woźniakowski [87], and the survey [65]. For example, it is known that adaptive information does not help (up to a factor of 2) for linear operator equations and the worst case error with respect to the unit ball of a normed space *F*. If *F* is only quasi-normed then the proofs must be modified, with a possible change of the constant 2. Nevertheless nonadaptive information is almost as good as adaptive information.

How much information is needed about the right-hand side $f \in F$ in order that we can solve Eq. (1) with an error $\varepsilon$? This question is answered by the minimal radii of information $r_n(S, F, H)$ (or the closely related Gelfand numbers). These numbers are a good measure for the *information complexity* of the operator equation. In contrast, the *output complexity* of the problem is measured by the nonlinear widths $e^{\mathrm{non}}_{n,C}(S, F, H)$. These numbers measure the cost of just outputting the approximation (with respect to an optimal basis $\mathcal{B} \in \mathcal{B}_C$). It is quite remarkable that, under general conditions, we obtain the same order

$$r_n(S, F, H) \asymp d^n(S, F, H) \asymp e^{\mathrm{non}}_{n,C}(S, F, H) \asymp n^{-t/d},$$

see Theorems 3 and 4.

Now we discuss *adaptive numerical schemes* for the numerical treatment of elliptic partial differential equations. Usually, these operator equations are solved by a Galerkin scheme, i.e., one defines an increasing sequence of finite-dimensional approximation spaces $G_{\Lambda_l} := \mathrm{span}\{\eta_\mu : \mu \in \Lambda_l\}$, where $G_{\Lambda_l} \subset G_{\Lambda_{l+1}}$, and projects the problem onto these spaces, i.e.,

$$\langle \mathcal{A} u_{\Lambda_l}, v \rangle = \langle f, v \rangle \quad \text{for all } v \in G_{\Lambda_l}.$$

To compute the actual Galerkin approximation, one has to solve a linear system

$$\mathbf{A}_{\Lambda_l} \mathbf{c}_{\Lambda_l} = \mathbf{f}_{\Lambda_l}, \quad \mathbf{A}_{\Lambda_l} = (\langle A\eta_{\mu'}, \eta_\mu \rangle)_{\mu,\mu' \in \Lambda_l}, \quad (\mathbf{f}_\Lambda)_\mu = \langle f, \eta_\mu \rangle, \quad \mu \in \Lambda_l.$$

Then the question arises how to choose the approximation spaces in a suitable way, since doing that in a somewhat clumsy fashion would yield huge linear systems and a very inefficient scheme. One natural way would be to use an updating strategy, i.e., one starts with a small set $\Lambda_0$, tries to estimate the (local) error, and only in regions where the error is large the index set is *refined*, i.e., further basis functions are added. Such an updating strategy is usually called an *adaptive numerical scheme* and it is characterized by the following facts: the sequence of approximation spaces is not a priori fixed but depends on the *unknown* solution $u$ of the operator equation, and the whole scheme should be self-regulating, i.e., it should work without a priori information on the solution. In principle, such an adaptive scheme consists of the following three steps:

| solve | estimate | refine |
|---|---|---|
| $\mathbf{A}_{\Lambda_l}\mathbf{c}_{\Lambda_l} = \mathbf{f}_{\Lambda_l}$ | $\Vert u - u_{\Lambda_l}\Vert = ?$<br>a posteriori<br>error estimator | add functions<br>if necessary. |

Note that the second step is highly nontrivial since the exact solution $u$ is unknown, so that clever a posteriori error estimators are needed. These error estimators should be local, since we want to refine (i.e., add basis functions) only in regions where the local error is large. Then another challenging task is to show that the refinement strategy leads to a convergent scheme and to estimate its order of convergence, if possible.

Recent developments indicate the promising potential of adaptive numerical schemes, see, e.g., [1,3–5,39,80,93] for finite element methods. However, to further explain the ideas and to make comparisons as simple as possible, we shall restrict ourselves to adaptive schemes based on wavelets. For simplicity, we shall mainly discuss the approach in [21]; for more sophisticated versions the reader is referred to [13–15,22]. The first step clearly must be the development of an a posteriori error estimator. Using the fact that $\mathcal{A}$ is boundedly invertible and the usual norm equivalences, comparing with (112), we obtain

$$\begin{aligned}
\Vert u - u_\Lambda\Vert_{H^s} &\asymp \Vert \mathcal{A}(u - u_\Lambda)\Vert_{H^{-s}} \\
&\asymp \Vert f - \mathcal{A}(u_\Lambda)\Vert_{H^{-s}} \\
&\asymp \Vert r_\Lambda\Vert_{H^{-s}} \\
&\asymp \left(\sum_{(j,\lambda)\in J\setminus\Lambda} 2^{-2sj}|\langle r_\Lambda, \psi_{j,\lambda}\rangle|^2\right)^{1/2} \\
&= \left(\sum_{(j,\lambda)\in J\setminus\Lambda} \delta_{j,\lambda}^2\right)^{1/2},
\end{aligned} \tag{37}$$

where the *residual weights* $\delta_{j,\lambda}$ can be computed as

$$\delta_{j,\lambda} = 2^{-sj}\left| f_{j,\lambda} - \sum_{(j',\lambda')\in\Lambda} \langle \mathcal{A}\psi_{j',\lambda'}, \psi_{j,\lambda}\rangle u_{j',\lambda'}\right| \quad \text{with } f_{j,\lambda} = \langle f, \psi_{j,\lambda}\rangle.$$

From (37), we observe that the sum of the residual weights gives rise to an efficient and reliable a posteriori error estimator. Each residual weight $\delta_{j,\lambda}$ can be interpreted as a local error indicator, so that the following natural refinement strategy suggests itself: Add wavelets in regions where the residual weights are large; that is, try to catch the bulk of the residual expansion in (37). Indeed, it can be shown that this strategy produces a convergent adaptive scheme, in principle. However, we are faced with a serious problem: the index set $J$ will not have finite cardinality, so that neither the error estimator nor the adaptive refinement strategy can be implemented. Nevertheless, there exist implementable variants, see again [13,21] for details. We start with the set

$$J_{j,\lambda,\varepsilon} : \{(j', \lambda') | \, |\langle \mathcal{A}\psi_{j',\lambda'}, \psi_{j,\lambda}\rangle| \ \varepsilon\text{-significant}\}$$

and define

$$a_{j,\lambda}(\Lambda, \varepsilon) := 2^{-sj} \left| \sum_{(j',\lambda')\in\Lambda\cap J_{j,\lambda,\varepsilon}} \langle \mathcal{A}\psi_{j',\lambda'}, \psi_{j,\lambda}\rangle u_{j',\lambda'} \right|.$$

(The expression '$\varepsilon$-significant' can be made precise by using the locality and the cancellation properties of a wavelet basis.) By employing the $a_{j,\lambda}(\Lambda, \varepsilon)$ we obtain another error estimator:

$$\|u - u_\Lambda\|_{H^s} \leqslant c \cdot \left( \left( \sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}^2 \right)^{1/2} + \varepsilon \|f\|_{H^{-s}} + \inf_{v\in\tilde{V}_\Lambda} \|F - v\|_{H^{-s}} \right).$$

Here $\tilde{V}_\Lambda$ denotes the approximation space spanned by the dual wavelets corresponding to $\Lambda$, see Appendix A.3 for details. Now, playing the same game for the $a_{j,\lambda}(\Lambda, \varepsilon)$ instead of the $\delta_{j,\lambda}$, we end up with a convergent and implementable adaptive strategy. To this end, the starting index set $\Lambda$ has to be determined such that $\inf_{v\in\tilde{V}_\Lambda} \|f - v\|_{H^{-s}} \leqslant c \cdot \mathrm{eps}$ and $\varepsilon(f, \mathrm{eps}, \theta)$ has to be computed. Then, there exists a constant $\kappa \in (0, 1)$ such that whenever $\tilde{\Lambda} \subset J,\ \Lambda \subset \tilde{\Lambda}$ is chosen so that

$$\left( \sum_{(j,\lambda)\in\tilde{\Lambda}\setminus\Lambda} a_{j,\lambda}(\Lambda, \varepsilon)^2 \right)^{1/2} \geqslant (1 - \theta) \left( \sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}(\Lambda, \varepsilon)^2 \right)^{1/2} \tag{38}$$

either

$$\|u - u_{\tilde{\Lambda}}\| \leqslant \kappa \|u - u_\Lambda\|, \quad \kappa \in (0, 1) \tag{39}$$

or

$$\left( \sum_{(j,\lambda)\in J\setminus\Lambda} a_{j,\lambda}(\Lambda, \varepsilon)^2 \right)^{1/2} \leqslant \mathrm{eps} \tag{40}$$

which implies that

$$\|u - u_\Lambda\| \leqslant \mathrm{eps} \cdot c. \tag{41}$$

For the proof and further details, the reader is again referred to [21].

**Remark 15.** (i) In order to avoid unnecessary technical and notational difficulties, we have not presented the explicit form of the function $\varepsilon(f, \mathrm{eps}, \theta)$. It depends in a complicated, but nevertheless computable way on the final accuracy eps, the control parameter $\theta$ , the $H^{-s}$-norm of the right-hand side *f*, and on the stability and ellipticity constants of the problem. For details, we refer again to [21].

(ii) The norm $\|\cdot\|$ in (39) and (41) clearly denotes the energy norm $\|v\| := \langle \mathcal{A}v, v\rangle$, which is equivalent to the Sobolev norm $H^s$, see again [50] for details.

(iii) Eqs. (39)–(41) obviously imply that the adaptive strategy in (38) converges. Indeed, the error is reduced by a factor of $\kappa$ at each step until the sum of the significant coefficients in (40) is smaller than the final accuracy, which by (41) means that the same property holds for the current Galerkin approximation.

(iv) Although the sum in the right-hand side of (38) formally still contains infinitely many coefficients, it can be checked that this sum in fact runs over a finite set, so that the adaptive strategy is implementable.

Let us now compare this concept of adaptivity with the notion of adaptive information explained above:

- From the discussion presented above, we have seen that adaptive wavelet schemes are not performed by gaining more and more information from the right-hand side *f* in an adaptive fashion. Instead they use the *residual* which depends on the right-hand side, the operator, and the domain. Moreover, we see that the starting index set $\Lambda$ is determined by the wavelet expansion of the right-hand side. That is, $\Lambda$ is given by some kind of best *n*-term approximation of *f*, which is assumed to be available or to be easily computable. In this sense, the adaptive wavelet schemes require *nonlinear information* about the problem.
- In the wavelet setting, the benchmark for the performance is the approximation order of the best *n*-term approximation of the solution, i.e., the numbers

$$\sup_{\|f\|_F \leqslant 1} \sigma_n(\mathcal{A}^{-1}f, \mathcal{B})_H. \tag{42}$$

  It has been shown quite recently in [13] that a judicious variant of the algorithm outlined above gives rise to the same order of approximation as best *n*-term approximation, while the number of arithmetic operations that are needed stays proportional to the number of unknowns. Here the authors implicitly assume that certain subroutines for fast matrix-vector multiplications, approximations of the right-hand sides and for thresholding are available, and that all these routines have to realize a given approximation rate. Moreover, it is assumed that the solution $u$ is contained in some Besov space $B^\alpha_p(L_p(\Omega))$, and hence $F$ is a suitable subset of $\mathcal{A}(B^\alpha_p(L_p(\Omega)))$, i.e., the admissible class of right-hand sides depends on the operator $\mathcal{A}$. Observe that, for given $F$ and $\mathcal{B}$, the numbers $e^{\mathrm{non}}_{n,C}(S, F, H)$ might be much smaller than the numbers in (42) since it is, in general, not clear whether a wavelet basis is optimal.
- The performance of an adaptive scheme is not compared with an *arbitrary linear* scheme. The reason for that is simple, and has already been explained earlier. It is indeed true that linear approximation often produces the same order as nonlinear (best *n*-term) approximations, see Theorems 2 and 4. However, for nonregular problems, it would be necessary to precompute the optimal basis $S(g_i)$ in advance, which is mostly too expensive and should be avoided in practice, see [24] for further details. One usually compares adaptive schemes with *uniform* methods for then a precomputation is not necessary. Therefore the use of an adaptive wavelet scheme is justified if it performs better than any uniform scheme. It is known that the order of approximation of uniform schemes is determined by the Sobolev regularity $H^t(\Omega)$ of the object we want to approximate whereas the approximation order of best *n*-term approximation

depends on the regularity in the specific Besov scale $B^t_\tau(L_\tau(\Omega))$, where

$$\frac{1}{\tau} = \frac{t-s}{d} + \frac{1}{2},$$

see [20,29] for details. Therefore adaptive schemes are justified if the Besov regularity of the exact solution is higher than its Sobolev regularity. For elliptic boundary value problems, there exist now many results in this direction, see, e.g., [16–19,23].

- In approximation theory, an approximation scheme that comes from a sequence of linear spaces that are uniformly refined is also called *linear approximation scheme*, which sometimes causes misunderstandings because these schemes are only special cases of the linear schemes considered, e.g., in Theorem 4. To avoid this confusion, we used the term uniform methods instead of linear methods.

**Remark 16.** In this paper we study the complexity of solving elliptic partial differential equations. We only deal with the deterministic setting. The randomized setting, where also the use of random numbers is allowed, is studied by Heinrich [51]. The complexity of solving elliptic PDE in the quantum model of computation (where one can use a certain nonclassical randomness) is studied in [52].

## 4. Proofs

### 4.1. Properties of widths

**Proof of Lemma 1.** *Step* 1: Part (i) is proved in [86] for the case where $F$ is normed. The general case is similar.

*Step* 2: To prove part (ii), we assume that $S(\{\|f\|_F \leqslant 1\})$ contains an $(n+1)$-dimensional ball $B \subset H$ of radius $r$ and that $N_n : F \to \mathbb{R}^n$ is continuous. Since $S^{-1}(B)$ is an $(n+1)$-dimensional bounded symmetric neighborhood of 0, it follows from the Borsuk Antipodality Theorem, see [28, paragraph 4], that there exists an $f \in \partial S^{-1}(B)$ with $N_n(f) = N_n(-f)$ and hence

$$S_n(f) = \varphi_n(N_n(f)) = \varphi_n(N_n(-f)) = S_n(-f)$$

for any mapping $\varphi_n : \mathbb{R}^n \to G$. Observe that $\|f\|_F = 1$. Because of $\|S(f) - S(-f)\| = 2r$ and $S_n(f) = S_n(-f)$ we obtain that the maximal error of $S_n$ on $\{\pm f\}$ is at least $r$. This proves

$$b_n(S, F, H) \leqslant e_n^{\text{cont}}(S, F, H).$$

Since we did not use the continuity of $\varphi_n$ also $b_n(S, F, H) \leqslant d^n(S, F, H)$ follows. □

**Proof of Lemma 3.** *Step* 1: Proof of (i). A corresponding assertion with $X$ and $Y$ normed linear spaces has been proved in [40]. This proof carries over without changes.

*Step* 2: Proof of (25). Let $\mathcal{B} = \{h_1, h_2, \ldots\}$ be a Riesz basis of $G$ with Riesz constants $A, B > 0$. Let this basis $\mathcal{B}$ and a corresponding mapping $S_n$ be optimal with respect to $I, F, G$ (up to some $\varepsilon > 0$ if necessary). Then the image of $\mathcal{B}$ under the mapping $S$ is a Riesz basis of $H$ with Riesz constants $A' = A/\|S^{-1}\|$ and $B' = B\,\|S\|$. From

$$\|Sf - (S \circ S_n)f\|_H \leqslant \|S\|\,\|f - S_n(f)\|_G$$

it follows that

$$e^{\mathrm{non}}_{n,C\,\|S^{-1}\|\,\|S\|}(S,F,H)\leqslant \|S\|\,e^{\mathrm{non}}_{n,C}(I,F,G)\,.$$

Replacing $C$ by $C/(\|S^{-1}\|\,\|S\|)$, the right-hand side in (25) follows.

Now, let $\mathcal{B}\subset H$ be a Riesz basis with Riesz constants $A, B>0$. Let $\mathcal{B}$ and a corresponding $S_n$ be optimal with respect to $S, F, H$ (again up to some $\varepsilon>0$ if necessary). From

$$\|If-(S^{-1}\circ S_n)f\|_G\leqslant \|S^{-1}\|\,\|Sf-S_n(f)\|_H$$

it follows that

$$e^{\mathrm{non}}_{n,C\,\|S^{-1}\|\,\|S\|}(I,F,G)\leqslant \|S^{-1}\|\,e^{\mathrm{non}}_{n,C}(S,F,H).$$

The proof of (24) follows from (27). □

Next we turn to the proof of Theorem 1. It is convenient for us to start with a simplified situation. For this we assume that $K\subset H$ is compact. We define

$$e^{\mathrm{non}}_{n,C}(K,H)=\inf_{\mathcal{B}\in\mathcal{B}_C}\sup_{u\in K}\sigma(u,\mathcal{B}) \tag{43}$$

and

$$e^{\mathrm{cont}}_{n}(K,H)=\inf_{N_n,\varphi_n}\sup_{u\in K}\|\varphi_n(N_n(u))-u\|, \tag{44}$$

where the infimum runs over all continuous mappings $\varphi_n:\mathbb{R}^n\to H$ and $N_n:K\to\mathbb{R}^n$. We prove the following result.

**Proposition 1.** *Let $K\subset H$ be compact. Then*

$$e^{\mathrm{cont}}_{4n+1}(K,H)\leqslant 2C\,e^{\mathrm{non}}_{n,C}(K,H). \tag{45}$$

**Proof.** Let $\mathcal{B}\in\mathcal{B}_C$ be given. Since $K$ is compact, we only need finitely many elements of $\mathcal{B}$, in the sense that

$$\sup_{u\in K}\|u-L_N(u)\|\leqslant\varepsilon \tag{46}$$

for

$$L_N(u)=\sum_{j=1}^{N}a_jh_j. \tag{47}$$

Here $L_N$ is the orthogonal projection onto the space that is generated by $h_1,\ldots,h_N$. The functionals $a_j$ are linear and continuous. Moreover, we know that

$$A\left(\sum_{j=1}^{N}|\alpha_j|^2\right)^{1/2}\leqslant\left\|\sum_{j=1}^{N}\alpha_jh_j\right\|\leqslant B\left(\sum_{j=1}^{N}|\alpha_j|^2\right)^{1/2} \tag{48}$$

with $B/A \leqslant C$. We may assume that $A = 1$. For a suitable $\mathcal{B} \in \mathcal{B}_C$ we obtain

$$\sup_{u \in K} \left\| \sum_{k=1}^{n} c_k \, h_{i_k} - L_N(u) \right\| \leqslant e_{n,C}^{\text{non}}(K, H) + \varepsilon. \tag{49}$$

Let $\beta > 0$. We define a modification of $L_N$ by

$$L_N^*(u) = \sum_{j=1}^{N} a_j^* h_j \tag{50}$$

where $a_j^* = a_j$ if $|a_j| \geqslant 2\beta$ and $a_j^* = 0$ if $|a_j| \leqslant \beta$. To make the $a_j^*$ continuous we define

$$a_j^* = 2\,\text{sgn}(a_j) \cdot (|a_j| - \beta)$$

for $|a_j| \in (\beta, 2\beta)$. We prove certain statements about $L_N^*$ and denote the best $n$-term approximation of $u$ by $u_n$.

Assume that for $u \in K$, there are $m > n$ of the $a_j$, see (47), such that $|a_j| \geqslant \beta$. Then we obtain

$$\|u_n - L_N(u)\| \geqslant (m - n)^{1/2} \beta$$

and with (49) we obtain

$$m - n \leqslant \frac{1}{\beta^2} (e_{n,C}^{\text{non}}(K, H) + \varepsilon)^2. \tag{51}$$

Now we consider the sum $\sum_{|a_j| < \beta} a_j^2$ for $u \in K$. We distinguish between those $j$ that are used for $u_n$ (there are only $n$ of those $j$) and the other indices and obtain

$$\sum_{|a_j|^2 < \beta} a_j^2 \leqslant n\beta^2 + (e_{n,C}^{\text{non}}(K, H) + \varepsilon)^2.$$

Now we are ready to estimate $\|L_N^*(u) - L_N(u)\|$ for $u \in K$. Observe that $|a_j^* - a_j| \leqslant \beta$ for any $j$. We obtain

$$\|L_N^*(u) - L_N(u)\| \leqslant B(m\beta^2 + n\beta^2 + (e_{n,C}^{\text{non}}(K, H) + \varepsilon)^2)^{1/2}.$$

Using the estimate (51) for $m$, we obtain

$$\|L_N^*(u) - L_N(u)\| \leqslant B(2n\beta^2 + 2(e_{n,C}^{\text{non}}(K, H) + \varepsilon)^2)^{1/2}.$$

Now we define $\beta$ by

$$n\beta^2 = (e_{n,C}^{\text{non}}(K, H) + \varepsilon)^2$$

and obtain the final error estimate (where we replace, for general $A$, the number $B$ by $B/A$)

$$\|L_N^*(u) - L_N(u)\| \leqslant \frac{2B}{A} \, (e_{n,C}^{\text{non}}(K, H) + \varepsilon).$$

In addition we obtain

$$m \leqslant 2n$$

and therefore $L_N^*$ yields a continuous $2n$-term approximation of $u \in K$ with error at most

$$\sup_{u\in K} \|L_N^*(u) - u\| \leqslant \frac{2B}{A} \, (e^{\mathrm{non}}_{n,C}(K, H) + \varepsilon) + \varepsilon.$$

The mapping $L_N^*$ is continuous and the image is a complex of dimension $2n$, see, e.g., [32]. Hence we have an upper bound for the so-called *Aleksandrov widths*, see [32,79]. By the famous theorem of Nöbeling, any such mapping can be factorized as $L_N^* = \varphi_{4n+1} \circ N_{4n+1}$ where $N_{4n+1} : K \to \mathbb{R}^{4n+1}$ and $\varphi_{4n+1} : \mathbb{R}^{4n+1} \to H$ are continuous. Hence the result is proved. $\square$

**Proof of Theorem 1.** The unit ball of $F$ is a compact subset of $G$ by assumption. From Proposition 1, we derive that

$$e^{\mathrm{cont}}_{4n+1}(I, F, G) \leqslant 2C \, e^{\mathrm{non}}_{n,C}(I, F, G).$$

Next we apply Lemma 3(ii), and obtain

$$e^{\mathrm{cont}}_{n}(S, F, H) \leqslant \|S\| \, e^{\mathrm{cont}}_{n}(I, F, G),$$

as well as

$$e^{\mathrm{non}}_{n,C}(I, F, G) \leqslant \|S^{-1}\| \, e^{\mathrm{non}}_{n,C/(\|S^{-1}\| \, \|S\|)}(S, F, H)\,.$$

Combining these inequalities, we are done. $\square$

*4.2. Widths of embeddings of weighted sequence spaces*

Having the wavelet characterization of Besov spaces in mind, cf. Appendices A.3 and A.4, we introduce the following scale of sequence spaces.

**Definition 3.** Let $0 < p, q \leqslant \infty$ and let $s \in \mathbb{R}$. Let $\nabla := (\nabla_j)_j$ be a sequence of subsets of finite cardinality of the set $\{1, 2, \dots, 2^d - 1\} \times \mathbb{Z}^d$. We suppose that there exist $0 < C_1 \leqslant C_2$ and $J \in \mathbb{N}$ such that the cardinality $|\nabla_j|$ of $\nabla_j$ satisfies

$$C_1 \leqslant 2^{-jd} \, |\nabla_j| \leqslant C_2 \quad \text{for all } j \geqslant J. \tag{52}$$

Then $b^s_{p,q}(\nabla)$, where $0 < q < \infty$, denotes the collection of all sequences $a = (a_{j,\lambda})_{j,\lambda}$ of complex numbers such that

$$\|a\|_{b^s_{p,q}} := \left( \sum_{j=0}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \left( \sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \right)^{q/p} \right)^{1/q} < \infty. \tag{53}$$

For $q = \infty$, we use the usual modification

$$\|a\|_{b^s_{p,\infty}} := \sup_{j=1,2,\dots} \; 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))} \left( \sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \right)^{1/p} < \infty. \tag{54}$$

If there is no danger of confusion we shall write $b^s_{p,q}$ instead of $b^s_{p,q}(\nabla)$.

**Remark 17.** In what follows, we shall let $e_{j,\lambda}$ denote the elements of the canonical orthonormal basis of $b^0_{2,2}$. Let $\sigma \in \mathbb{R}$. It is obvious that the linear mapping $L_\sigma$ defined by

$$L_\sigma \, e_{j,\lambda} := 2^{-\sigma j} \, e_{j,\lambda} \quad \text{for all } j, \ \lambda,$$

extends to an isomorphism from $b^s_{p,q}$ onto $b^{s+\sigma}_{p,q}$ (simultaneously for all $s, p, q$) with $\|L_\sigma\| = 1$.

In the framework of these sequence spaces it is very easy to prove embedding theorems, cf. [57].

**Lemma 4.** *Let* $0 < p_0, p_1, q_0, q_1 \leqslant \infty$, $s \in \mathbb{R}$, *and* $t \geqslant 0$.

(i) *The embedding*

$$b^{s+t}_{p_0,q_0}(\nabla) \hookrightarrow b^s_{p_1,q_1}(\nabla)$$

*exists* (*as a set theoretic inclusion*) *if and only if it is continuous if and only if either*

$$t > d\left(\frac{1}{p_0} - \frac{1}{p_1}\right)_+ \tag{55}$$

*or*

$$t = d\left(\frac{1}{p_0} - \frac{1}{p_1}\right)_+ \quad \textit{and} \quad q_0 \leqslant q_1.$$

(ii) *The embedding*

$$b^{s+t}_{p_0,q_0}(\nabla) \hookrightarrow b^s_{p_1,q_1}(\nabla)$$

*is compact if and only if* (55) *holds.*

The main result of this subsection consists in the following:

**Theorem 6.** *Let* $0 < p, p_0, p_1 \leqslant \infty$, $0 < q, q_0, q_1 \leqslant \infty$, *and* $s \in \mathbb{R}$.

(i) *Suppose that*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+ \tag{56}$$

*holds. Then, for any* $C \geqslant 1$, *we have*

$$e^{\mathrm{non}}_{n,C}(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp n^{t/d}.$$

(ii) *Suppose that* (56) *holds. Then we have*

$$e^{\mathrm{lin}}_n(I, b^{s+t}_{p,q}, b^s_{2,2}) \asymp \begin{cases} n^{-t/d} & \textit{if } 2 \leqslant p \leqslant \infty, \\ n^{-t/d+1/p-1/2} & \textit{if } 0 < p < 2. \end{cases}$$

(iii) *Suppose that* (55) *holds. Then we have*

$$e^{\mathrm{cont}}_n(I, b^{s+t}_{p_0,q_0}, b^s_{p_1,q_1}) \asymp n^{-t/d}.$$

**Remark 18.** In part (i) there is an interesting limiting case. Suppose $0 < p < 2$ and $t = d(1/p - 1/2)$. Then the embedding $b^{s+t}_{p,p} \hookrightarrow b^s_{2,2}$ exists, cf. Lemma 4, and

$$\left(\sum_{n=1}^{\infty} \left[n^{t/d}\, \sigma_n(a, \mathcal{B})_{b^s_{2,2}}\right]^p \frac{1}{n}\right)^{1/p} < \infty \quad \text{if and only if } a \in b^{s+t}_{p,p}.$$

In view of Lemma 4(ii), this shows that $\lim_{n\to\infty} e^{\rm non}_{n,C}(S, F, H) = 0$ does not imply compactness of $S$.

The proof of Theorem 6 requires some preparations. It will be given in Sections 4.2.2–4.2.4.

*4.2.1. The Bernstein widths of the identity operator*

We concentrate on the estimate from below. For later use we treat a more general situation.

**Lemma 5.** *Let* $0 < p_0, p_1, q_0, q_1 \leqslant \infty$, $s \in \mathbb{R}$ *and* $t > 0$ *such that* (55) *holds. Then there exists a positive constant c such that*

$$b_n(I, b^{s+t}_{p_0,q_0}, b^s_{p_1,q_1}) \geqslant c \begin{cases} n^{-t/d} & \text{if } 0 < p_0 \leqslant p_1 \leqslant \infty, \\ n^{-t/d+1/p_0-1/p_1} & \text{if } 0 < p_1 < p_0 \leqslant \infty, \end{cases} \tag{57}$$

*holds for all n.*

**Proof.** The Bernstein numbers are monotonic in $n$. So it will be enough to prove the assertion for sufficiently large $n$. Consequently, we may assume that there is a natural number $N \geqslant J$, as well as positive constants $c_1$ and $c_2$, such that

$$c_1 2^{Nd} \leqslant n \leqslant c_2 2^{Nd}.$$

*Step* 1: Let $0 < p_0 \leqslant p_1$. Using Hölder's inequality we find

$$\begin{aligned}\left\| \sum_{\lambda\in\nabla_N} b_\lambda e_{N,\lambda} \,\middle|\, b^{s+t}_{p_0,q_0} \right\| &= 2^{N(s+t+d/2-d/p_0)} \left(\sum_{\lambda\in\nabla_N} |b_\lambda|^{p_0}\right)^{1/p_0} \\ &\leqslant 2^{N(s+t+d/2-d/p_0)}\, |\nabla_N|^{1/p_0-1/p_1} \left(\sum_{\lambda\in\nabla_N} |b_\lambda|^{p_1}\right)^{1/p_1} \\ &\leqslant C_2 2^{Nt} \left\| \sum_{\lambda\in\nabla_N} b_\lambda\, e_{N,\lambda} \,\middle|\, b^s_{p_1,q_1} \right\| \\ &\leqslant c_3 n^{t/d} \left\| \sum_{\lambda\in\nabla_N} b_\lambda\, e_{N,\lambda} \,\middle|\, b^s_{p_1,q_1} \right\|,\end{aligned}$$

where $C_2$ corresponds to (52). Consequently, the unit ball in $b^{s+t}_{p_0,q_0}$ contains the $n$-dimensional ball (spanned by the vectors $e_{N,\lambda}$, $\lambda \in \nabla_N$) with radius $c_3^{-1}\, n^{-t/d}$. This proves

$$b_n(I, b^{s+t}_{p_0,q_0}, b^s_{p_1,q_1}) \geqslant c\, n^{-t/d}$$

for some positive constant $c$ independent of $n$.

*Step* 2: If $p_0 > p_1$, then Hölder's inequality (used in the second line of the estimate in Step 1) will be replaced by the monotonicity of the $\ell_r$-norms and we obtain

$$\begin{aligned}\left\| \sum_{\lambda\in\nabla_N} b_\lambda e_{N,\lambda} \middle| b^{s+t}_{p_0,q_0} \right\| &= 2^{N(s+t+d/2-d/p_0)} \left( \sum_{\lambda\in\nabla_N} |b_\lambda|^{p_0} \right)^{1/p_0} \\ &\leqslant 2^{N(s+t+d/2-d/p_0)} \left( \sum_{\lambda\in\nabla_N} |b_\lambda|^{p_1} \right)^{1/p_1} \\ &\leqslant c_5 2^{N(t+d/p_1-d/p_0)} \left\| \sum_{\lambda\in\nabla_N} b_\lambda\, e_{N,\lambda} \middle| b^{s}_{p_1,q_1} \right\| .\end{aligned}$$

This time the unit ball in $b^{s+t}_{p_0,q_0}$ contains the $n$-dimensional ball with radius

$$c_5^{-1} 2^{-N(t+d/p_1-d/p_0)}.$$

This proves our claims. $\square$

**Remark 19.** In the one-dimensional periodic situation, estimates of the Bernstein numbers from above are also known, due to Tsarkov and Maiorov, cf. [85, Theorem 12, p. 194]. Let $1 \leqslant p \leqslant \infty$ and $s > 0$. By $\mathring{W}^s_p$ we denote the collection of all $2\pi$-periodic functions $f$ with Weyl derivative of order $s$ belonging to $L_p(\mathbb{T})$ and satisfying $\int_{-\pi}^{\pi} f(x)\, dx = 0$. Then

$$b_n(I, \mathring{W}^t_{p_0}, L_{p_1}) \asymp \begin{cases} n^{-t} & \text{if } 1 \leqslant p_0 \leqslant p_1 \leqslant \infty \text{ or } 1 \leqslant p_1 \leqslant p_0 \leqslant 2 \text{ and } t > 0, \\ n^{-t+1/p_0-1/p_1} & \text{if } 2 \leqslant p_1 < p_0 \leqslant \infty \text{ and } t > 1/p_0, \\ n^{-t+1/p_0-1/2} & \text{if } 1 \leqslant p_1 \leqslant 2 \leqslant p_0 \leqslant \infty \text{ and } t > 1/p_0. \end{cases}$$

This should be compared with Lemma 5 for $s = 0$ and $d = 1$.

### *4.2.2. Best n-term approximation in the framework of sequence spaces*

We prepare the proof of part (i) of Theorem 6. Also, here we treat a more general situation. Let $\mathcal{B}$ denote the canonical basis $(e_{j,\lambda})_{j,\lambda}$ in $b^0_{2,2}(\nabla)$. Then our aim in this subsection consists in a characterization of the behavior of the best $n$-term approximation of a given element $a \in b^{s+t}_{p_0,q_0}$ with respect to $\mathcal{B}$.

The main result of this subsection reads as follows:

**Theorem 7.** *Let* $0 < p_0, p_1, q_0, q_1 \leqslant \infty$, $s \in \mathbb{R}$ *and* $t > 0$ *such that* (55) *holds. Then we have*

$$\sup \left\{ \sigma_n(a, \mathcal{B})_{b^s_{p_1,q_1}} \; : \; \| a \|_{b^{s+t}_{p_0,q_0}} \leqslant 1 \right\} \asymp n^{-t/d}. \tag{58}$$

We start with some preparations. Let $U$ denote the unit ball in $b^{s+t}_{p_0,\infty}$. Then

$$a = \sum_{j=0}^{\infty} \sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda} \quad \text{and} \quad \sup_{j=0,1,\dots} 2^{j(s+t+d(1/2-1/p_0))} \left( \sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^{p_0} \right)^{1/p_0} \leqslant 1.$$

The following lemma will be of some use:

**Lemma 6.** *Let* $0 < p_0 \leqslant p_1$ *and suppose that*

$$t > d\left(\frac{1}{p_0} - \frac{1}{p_1}\right). \tag{59}$$

*For all* $a \in U$ *and all* $n \geqslant 1$ *there exists a natural number* $K := K(a, n)$ *such that*

$$\left\| a - \sum_{j=0}^{K} \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda} \middle| b^s_{p_1,q_1} \right\| \leqslant n^{-t/d}$$

*holds.*

**Proof.** We define

$$T_j := \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda}, \quad j = 0, 1 \ldots .$$

Then one has

$$a - \sum_{j=0}^{K} \sum_{\lambda \in \nabla_j} a_{j,\lambda}\, e_{j,\lambda} = \sum_{j>K} T_j .$$

Since $0 < p_0 \leqslant p_1 \leqslant \infty$, the monotonicity of the $\ell_q$-norms and $a \in U$ lead to

$$\begin{aligned} \| T_j \,|b^s_{p_1,q_1}\| &\leqslant 2^{j(s+d/2-d/p_1)} \left( \sum_{\lambda \in \nabla_j} |a_{j,\lambda}|^{p_0} \right)^{1/p_0} \\ &\leqslant 2^{-j(t+d(1/p_0-1/p_1))}. \end{aligned}$$

Let $u = \min(1, p_1, q_1)$. Consequently, using (59) and choosing $K$ large enough, we find

$$\begin{aligned} \left\| \sum_{j \geqslant K} T_j \middle| b^s_{p_1,q_1} \right\|^u &\leqslant \sum_{j \geqslant K} \| T_j \,|b^s_{p_1,q_1}\|^u \leqslant \sum_{j \geqslant K} 2^{-ju\left[t+d(1/p_0-1/p_1)\right]} \\ &\leqslant c 2^{-Ku(t+d(1/p_0-1/p_1))} \leqslant n^{-tu/d}. \end{aligned}$$

This proves the claim. □

The basic step in deriving an upper estimate of $\sigma_n(a, \mathcal{B})$ is the following proposition. Again $U$ denotes the unit ball in $b^{s+t}_{p_0,\infty}$.

**Proposition 2.** *Let* $0 < p_0 \leqslant p_1 \leqslant \infty$. *Let* $a \in U$, $n \in \mathbb{N}$, *and let* $K = K(a, n)$ *be as in Lemma* 6. *Then there exists an approximation*

$$S_n a := \sum_{j=0}^{K} \sum_{\lambda \in \nabla_j} a^*_{j,\lambda}\, e_{j,\lambda} \tag{60}$$

*of $a$, which satisfies the following*:

(i) *The coefficients $a^*_{j,\lambda}$ depend continuously on $a$.*
(ii) *The number of nonvanishing entries is bounded by $c \cdot n$.*
(iii) $\|a - S_n a | b^s_{p_1,q_1}\| \leqslant c n^{-t/d}$, $n = 1, 2, \ldots$ .

*Here $c$ can be chosen independently of $a$ and $n$.*

**Proof.** Observe that it will be enough to prove the claim for natural numbers $n = 2^{Nd}$, where $N \in \mathbb{N}$. We define

$$\delta := \frac{t - d\ (1/p_0 - 1/p_1)}{2\,(1/p_0 - 1/p_1)},$$

$$\varepsilon_j := \begin{cases} 0 & \text{if } 1 \leqslant j \leqslant N \\ n^{-1/p_0} 2^{-jd(1/2-1/p_0)} 2^{-jt} 2^{(j-N)\delta/p_0} & \text{if } j > N, \end{cases} \tag{61}$$

$$\Lambda^*_j := \left\{ \lambda \in \nabla_j \ : \ |a_{j,\lambda}|\, 2^{sj} \geqslant \varepsilon_j \right\}, \qquad j = 0, 1, \ldots \ . \tag{62}$$

Then, if $j > N$,

$$\begin{aligned} |\Lambda^*_j| &= \sum_{\lambda \in \Lambda^*_j} 1 \leqslant \sum_{\lambda \in \Lambda^*_j} 2^{jsp_0} \frac{|a_{j,\lambda}|^{p_0}}{\varepsilon_j^{p_0}} \\ &\leqslant \sum_{\lambda \in \nabla_j} n 2^{jd(1/2-1/p_0)p_0} 2^{jtp_0} 2^{-(j-N)\delta} 2^{jsp_0} |a_{j,\lambda}|^{p_0} \\ &= n 2^{-(j-N)\delta} \sum_{\lambda \in \nabla_j} 2^{j(s+t+d(1/2-1/p_0))p_0}\, |a_{j,\lambda}|^{p_0} \\ &\leqslant n 2^{-(j-N)\delta}\, \|\, a\, | b^{s+t}_{p_0,\infty} \|^{p_0} \\ &\leqslant n 2^{-(j-N)\delta}. \end{aligned} \tag{63}$$

Now a typical method to approximate $a$ would be to choose $a^*_{j,\lambda} = a_{j,\lambda}$, $j \in \Lambda^*_j$, and zero otherwise. However, this selection does not depend continuously on $a$. Therefore we use the following variant. Let $g_j$ denote the following piecewise linear and odd function:

$$g_j(x) := \begin{cases} 0 & \text{if } 0 \leqslant x \leqslant 2^{-js} \varepsilon_j, \\ x & \text{if } x \geqslant 2 \cdot 2^{-js} \varepsilon_j, \\ \text{linear} & \text{if } x \in (2^{-js}\varepsilon_j, 2 \cdot 2^{-js}\varepsilon_j)\,. \end{cases} \tag{64}$$

Then we set

$$a^*_{j,\lambda} := g_j(a_{j,\lambda}) \tag{65}$$

and consider the associated approximation (60). Let us prove that $S_n$ will do the job.

*Step* 1: We shall prove (i) and (ii). Observe

$$\left| \bigcup_{j=0}^{K} \Lambda^*_j \right| \leqslant c_1 \sum_{j=0}^{N} 2^{jd} + \sum_{j=N+1}^{K} n 2^{-(j-N)\delta} \leqslant c_2\, n,$$

cf. (63). The constant $c_2$ is independent of $a$, $K$, and $n$. This proves (i) and (ii).

*Step* 2: Proof of (iii). We have

$$a - S_n a \;=\; a - \sum_{j=0}^{K} \sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda} \;+\; \sum_{j=0}^{K} T_j^* \;:=\; \Sigma_1 + \Sigma_2,$$

where

$$T_j^* \;=\; \sum_{\lambda\in\nabla_j} \left(a_{j,\lambda} - a_{j,\lambda}^*\right) e_{j,\lambda}.$$

From Lemma 6, we can conclude that $\|\Sigma_1|b^s_{p_1,q_1}\| \leqslant n^{-t/d}$ for $K$ large enough. Therefore it remains to estimate $\|T_j^*\,|b^s_{p_1,q_1}\|$. Since $|g_j(x) - x| \leqslant |x|$ and $a^*_{j,\lambda} = a_{j,\lambda}$ for $|a_{j,\lambda}| \geqslant 2\varepsilon_j 2^{-js}$, we obtain

$$\begin{aligned}
|\,a_{j,\lambda} - a^*_{j,\lambda}\,|^{p_1} &\leqslant |a_{j,\lambda}|^{p_1} \\
&\leqslant |a_{j,\lambda}|^{p_0} |a_{j,\lambda}|^{p_1-p_0} \\
&\leqslant |a_{j,\lambda}|^{p_0} (2\varepsilon_j)^{p_1-p_0} 2^{-js(p_1-p_0)}\,.
\end{aligned}$$

This will be used to estimate the norm of $T_j^*$ as follows:

$$\begin{aligned}
\|\,T_j^*\,|b^s_{p_1,q_1}\| &= 2^{j(s+d(1/2-1/p_1))} \Bigg( \sum_{k\in\nabla_j} |a_{j,\lambda} - a^*_{j,\lambda}|^{p_1} \Bigg)^{1/p_1} \\
&\leqslant c_1 2^{jd(1/2-1/p_1)}\, 2^{jsp_0/p_1} \varepsilon_j^{1-p_0/p_1} \Bigg( \sum_{k\in\nabla_j} |a_{j,\lambda}|^{p_0} \Bigg)^{1/p_1} \\
&\leqslant c_1 \varepsilon_j^{1-p_0/p_1} 2^{jd/2} 2^{-jtp_0/p_1} 2^{-jdp_0/(2p_1)} \\
&\quad \times \Bigg( \sum_{\lambda\in\nabla_j} 2^{j(s+t+d(1/2-1/p_0))p_0} |a_{j,\lambda}|^{p_0} \Bigg)^{1/p_1} \\
&\leqslant c_2 \varepsilon_j^{1-p_0/p_1} 2^{-j(t+d/2-dp_1/(2p_0))p_0/p_1}\ \|a\,|b^{s+t}_{p_0,\infty}\|^{p_0/p_1} \\
&\leqslant c_2 \varepsilon_j^{1-p_0/p_1} 2^{-j(t+d/2-dp_1/(2p_0))p_0/p_1},
\end{aligned}$$

where again $c_2$ does not depend on $a$ and $n$. For $j > N$ we continue by employing the concrete value of $\varepsilon_j$ and obtain

$$\begin{aligned}
\|T_j^*|b^s_{p_1,q_1}\| &\leqslant c_2 \left( n^{-1/p_0} 2^{-jd(1/2-1/p_0)} 2^{-jt} 2^{(j-N)\delta/p_0} \right)^{1-p_0/p_1} \\
&\quad \times 2^{-j(t+d/2-dp_1/(2p_0))p_0/p_1} \\
&= c_2 n^{1/p_1-1/p_0}\, 2^{-N\delta(1/p_0-1/p_1)} 2^{-j(t-d(1/p_0-1/p_1)-\delta/p_0+\delta/p_1)}.
\end{aligned}$$

By construction $T_j^* = 0$ if $j \leqslant N$, by definition, we have

$$t - d\left(\frac{1}{p_0} - \frac{1}{p_1}\right) > \delta\left(\frac{1}{p_0} - \frac{1}{p_1}\right).$$

Hence, with $u = \min(1, p_1, q_1)$, we have

$$\|\Sigma_2\,|b^s_{p_1,q_1}\|^u \leqslant c_2^u \left(n^{1/p_1-1/p_0}\,2^{-N\delta(1/p_0-1/p_1)}\right)^u \sum_{j=N+1}^{K} 2^{-ju(t-d(1/p_0-1/p_1)-\delta/p_0+\delta/p_1)}$$
$$\leqslant c_3 \left(n^{1/p_1-1/p_0}\,2^{-N\delta(1/p_0-1/p_1)}\right)^u\, 2^{-Nu(t-d(1/p_0-1/p_1)-\delta/p_0+\delta/p_1)}$$
$$= c_3 \left(n^{1/p_1-1/p_0}\right)^u\, 2^{-Nu(t-d(1/p_0-1/p_1))},$$

with $c_3$ independent of $K$, $n$ and $a$. Recalling that $2^{Nd} = n$, we end up with

$$\|\Sigma_2\,|b^s_{p_1,q_1}\| \leqslant c_3 n^{-t/d}.$$

This finishes the proof of Proposition 2. $\square$

For completeness and better reference we formulate the counterpart of Proposition 2 in the case $p_0 \geqslant p_1$.

**Proposition 3.** *Let $0 < p_1 \leqslant p_0 \leqslant \infty$. Let $a \in U$ (the unit ball in $b^{s+t}_{p_0,\infty}$) and $2^{Nd} \leqslant n \leqslant 2^{(N+1)d}$. Then the approximation*

$$S_n a := \sum_{j=0}^{N} \sum_{\lambda\in\nabla_j} a_{j,\lambda}\, e_{j,\lambda} \tag{66}$$

*of a satisfies the following*:

(i) *The coefficients $a_{j,\lambda}$ depend continuously on a.*
(ii) *The number of nonvanishing entries is bounded by $c \cdot n$.*
(iii) $\|a - S_n a\,|b^s_{p_1,q_1}\| \leqslant c\, n^{-t/d}, \quad n = 1, 2, \ldots$ .

*Here, c can be chosen independently of a and n.*

**Proof.** The proof is elementary. $\square$

**Proof of Theorem 7.** The estimate from above follows from Propositions 2 and 3, as well as the continuous embedding $b^{s+t}_{p_0,q_0} \hookrightarrow b^{s+t}_{p_0,\infty}$. For the estimate from below, it will be enough to consider $n = 2^{Nd}$, where $N \geqslant J$ and $N \in \mathbb{N}$. Let $K$ be the smallest natural number such that $C_1 2^{Kd} \geqslant 2$ (here $C_1$ is the same constant as in (52)). Then

$$n \leqslant \frac{C_1 2^{(N+K)d}}{2} \leqslant \frac{1}{2}\,|\nabla_{N+K}|.$$

Let $\Gamma \subset \nabla_{N+K}$ with $|\Gamma| = n$. We define

$$a = |\nabla_{N+K}|^{-1/p_0}\, 2^{-(N+K)(s+t+d(1/2-1/p_0))} \sum_{\lambda\in\nabla_{N+K}} e_{N+K,\lambda}.$$

Consequently $\|a\|_{b^{s+t}_{p_0,q_0}} = 1$ for any $q_0$. Furthermore, we find

$$\|a - S_n a\|_{b^s_{p_1,q_1}} \geqslant \left\| \sum_{\lambda\in\nabla_{N+K}\setminus\Gamma} |\nabla_{N+K}|^{-1/p_0}\, 2^{-(N+K)(s+t+d(1/2-1/p_0))}\, e_{N+K,\lambda} \right\|_{b^s_{p_1,q_1}}$$

$$
\begin{aligned}
&= |\nabla_{N+K}|^{-1/p_0}\, 2^{-(N+K)(t+d(1/p_1-1/p_0))} |\nabla_{N+K} \setminus \Gamma|^{1/p_1} \\
&\geqslant \frac{C_1^{1/p_1}}{2^{1/p_1}\, C_2^{1/p_0}}\, 2^{-(N+K)t} \\
&= \frac{C_1^{1/p_1}}{2^{1/p_1}\, C_2^{1/p_0}}\, 2^{-Kt}\, n^{-t/d}
\end{aligned}
$$

(also $C_2$ has the same meaning as in (52)). It is clear that an optimal $\Gamma$ with $|\Gamma| = n$ has to be a subset of $\nabla_{N+K}$. This completes the proof of the estimate from below. $\square$

**Proof of Theorem 6(i).** The estimate from above is covered by Theorem 7; the estimate from below follows from Theorems 1 and 6(iii). $\square$

**Remark 20.** Stepanets [78] has investigated the quantities

$$\sigma_n(a, B)_{b^s_{p_1,q_1}}$$

for the specific case

$$s = d\left(\frac{1}{p_1} - \frac{1}{2}\right) \quad \text{with } p_1 = q_1.$$

In this special case, the associated nonlinear widths related to quite general smoothness spaces are studied. He proved explicit formulas from which the asymptotic behavior could be derived.

*4.2.3. The manifold widths of the identity*

**Proof of Theorem 6(iii).** Without loss of generality we may choose $s = 0$, cf. Lemma 3(ii) and Remark 17.

*Step* 1: The estimate from above. In the case $p_1 = q_1 = 2$ we may use Propositions 1–3 to get the desired inequality. However, for the general case we have to modify the argument. We follow the arguments used in [32]. Let $U$ denote the unit ball in $b^t_{p_0,q_0}$. As explained there Propositions 2 and 3 guarantee that

$$a^n(U, b^0_{p_1,q_1}) \leqslant c\, n^{-t/d},$$

where $a^n$ denotes the Alexandroff-co-width, cf. [32] for details. But

$$e^{\text{cont}}_{2n+1}(U, b^0_{p_1,q_1}) \leqslant a^n(U, b^0_{p_1,q_1}),$$

cf. [32,40]. Let us mention that in the literature quoted the target space was always a normed linear space. But the arguments carry over to quasi-normed linear spaces.

*Step* 2: The estimate from below. Lemmas 1 and 5 yield the lower estimate in case $0 < p_0 \leqslant p_1 \leqslant \infty$.

Now, let $p_1 < p_0 \leqslant \infty$. Let $\varepsilon > 0$. We consider the diagram

$$
\begin{array}{ccc}
b^0_{p_1,q_1} & \xrightarrow{I_3} & b^{-d(1/p_1-1/p_0)-\varepsilon}_{p_0,\infty} \\
I_2 \nwarrow & & \nearrow I_1 \\
& b^t_{p_0,q_0}, &
\end{array}
$$

where $I_1$, $I_2$ and $I_3$ are identity operators. Then (23) yields

$$e_{2n}^{\text{cont}}(I_1, b^t_{p_0,q_0}, b^{-d(1/p_1-1/p_0)-\varepsilon}_{p_0,\infty}) \leqslant e_n^{\text{cont}}(I_2, b^t_{p_0,q_0}, b^0_{p_1,q_1}) e_n^{\text{cont}}(I_3, b^0_{p_1,q_1}, b^{-d(1/p_1-1/p_0)-\varepsilon}_{p_0,\infty})$$

which implies that

$$c_1 n^{-t/d-1/p_1+1/p_0-\varepsilon/d} \leqslant c_2 e_n^{\text{cont}}(I_2, b^t_{p_0,q_0}, b^0_{p_1,q_1})\ n^{-1/p_1+1/p_0-\varepsilon/d}$$

for some positive $c_1$ and $c_2$ (independent of $n$), see Lemmas 5, 1, and Step 1. □

**Remark 21.** It is clear from the proof given above that the knowledge of the Bernstein widths is not enough to establish the estimate from below of $e_n^{\text{cont}}$. Here the multiplicativity of the numbers $e_n^{\text{cont}}$, cf. (23), is crucial. This seems to be overlooked in [32].

*4.2.4. The approximation numbers of the identity*

**Proof of Theorem 6(ii).** *Step* 1: Let $2 \leqslant p \leqslant \infty$. From Proposition 3 we obtain the estimate from above with $S_n$ given by (66). The estimate from below is covered by (58).

*Step* 2: Let $0 < p < 2$. Without loss of generality we assume $s = 0$. Let $S_n$ be defined by (66). The estimate from above is easily derived by using the monotonicity of the $\ell_r$-norms and $t + d(1/2 - 1/p) > 0$:

$$\begin{aligned}
\|a - S_n a\,|b^0_{2,2}\|^2 &\leqslant \sum_{j=N+1}^{\infty} \left( \sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \right)^{2/p} \\
&\leqslant \left( \sum_{j=N+1}^{\infty} 2^{-2j(t+d(1/2-1/p))} \right) \\
&\quad \times \left( \sup_{j\geqslant N+1} 2^{j(t+d(1/2-1/p))} \left( \sum_{\lambda\in\nabla_j} |a_{j,\lambda}|^p \right)^{1/p} \right)^2 \\
&\leqslant c 2^{-2N(t+d(1/2-1/p))} \|a|b^t_{p,\infty}\|^2 \\
&\leqslant c \left( n^{-t/d-1/2+1/p} \|a|b^t_{p,q}\| \right)^2,
\end{aligned}$$

where $c$ does not depend on $n$ and $a$. For the estimate from below, we use the obvious fact that the optimal approximation of an element in a Hilbert space is given by the partial sum with respect to an orthonormal basis. Hence, if $\widetilde{S}_n$ is a linear operator of rank at most $n$ then

$$\|a - \widetilde{S}_n a|b_{0,0}\| \geqslant \|a - S_n a|b_{0,0}\|,$$

where $S_n$ is defined by (66). We put

$$a := \sum_{j=0}^{N+1} e_{j,\lambda_j},$$

where $\lambda_j \in \nabla_j$ can be chosen arbitrarily. Then

$$\|a|b^t_{p,q}\| = \left(\sum_{j=0}^{N+1} 2^{j(t+d(1/2-1/p))q}\right)^{1/q} \geqslant 2^{N(t+d(1/2-1/p))}$$

for some positive $c$ independent of $n$ and

$$\|a - S_n a|b^0_{2,2}\| = 1.$$

This implies

$$\|I - S_n\,|b^t_{p,q}\| \geqslant \frac{1}{2^{N(t+d(1/2-1/p))}},$$

which finishes the proof of the lower bound. □

**Remark 22.** Notice that in any case, an order-optimal approximation is given by an appropriate partial sum, see (66).

*4.2.5. The Gelfand widths of the identity*

What we will do here relies on a result of Gluskin [45,46] about the Gelfand widths of the embedding $\ell^m_p \to \ell^m_2$ which we now recall. Let $1/p + 1/p' = 1$. For all natural numbers $m$ and $n$, where $n \leqslant m$, it holds that

$$d^n(I, \ell^m_p, \ell^m_2) \asymp \begin{cases} (m-n+1)^{\frac{1}{2}-\frac{1}{p}} & \text{if } 2 \leqslant p \leqslant \infty, \\ 1 & \text{if } 1 \leqslant p < 2 \text{ and } 1 \leqslant n \leqslant m^{2/p'}, \\ m^{1/p'} n^{-1/2} & \text{if } 1 \leqslant p < 2 \text{ and } m^{2/p'} \leqslant n \leqslant m. \end{cases} \tag{67}$$

A simple monotonicity argument leads to the following supplement to $p = 1$. There exists a constant $c$, independent of $m$ and $n$, such that

$$d^n(I, \ell^m_p, \ell^m_2) \leqslant c n^{-1/2} \tag{68}$$

if $0 < p < 1$ and $1 \leqslant n \leqslant m$.

The Gelfand widths are examples of so-called $s$-numbers, cf. [10,72,73]. Following [72, 2.2.4, p. 80] we associate with the sequence of Gelfand widths the following operator ideals. Let $F$ and $E$ be quasi-Banach spaces and denote by $\mathcal{L}(F, E)$ the class of all linear continuous operators $T : F \to E$. Then, for $0 < p < \infty$, we put

$$\mathcal{L}^{(c)}_{r,\infty} := \left\{ T \in \mathcal{L}(F, E) : \quad \sup_{n \in \mathbb{N}} n^{1/r} d^n(T) < \infty \right\}.$$

Equipped with the quasi-norm

$$\lambda_r(T) := \sup_{n \in \mathbb{N}} n^{1/r}\, d^n(T),$$

the set $\mathcal{L}^{(c)}_{r,\infty}$ becomes a quasi-Banach space. For such quasi-Banach spaces there always exist a real number $\varrho \in (0, 1]$ and an equivalent quasi-norm, here denoted by $\|\cdot|\mathcal{L}^{(c)}_{r,\infty}\|$, such that

$$\|T_1 + T_2 \,|\mathcal{L}_{r,\infty}^{(c)}\|^{\varrho} \leqslant \|T_1 \,|\mathcal{L}_{r,\infty}^{(c)}\|^{\varrho} + \|T_2 \,|\mathcal{L}_{r,\infty}^{(c)}\|^{\varrho} \tag{69}$$

holds for all $T_1, T_2 \in \mathcal{L}_{r,\infty}^{(c)}$.

To shorten notation we shall use the abbreviation $I_{p,q}^m$ for the identity $I : \ell_p^m \to \ell_q^m$. It is not complicated to check that (67), (68) imply the following estimates for $\|I_{p,2}^m \,|\mathcal{L}_{r,\infty}^{(c)}\|$, cf. [58].

**Lemma 7.** *Let* $0 < r < \infty$.

(i) *Let* $2 \leqslant p \leqslant \infty$. *Then*

$$\|I_{p,2}^m \,|\mathcal{L}_{r,\infty}^{(c)}\| \asymp m^{1/r-1/p+1/2} \tag{70}$$

*holds*.

(ii) *Let* $1 < p < 2$. *Then*

$$\|I_{p,2}^m \,|\mathcal{L}_{r,\infty}^{(c)}\| \asymp \begin{cases} m^{1/r-1/p+1/2} & \text{if } 0 < r \leqslant 2\,, \\ m^{2/(rp')} & \text{if } 2 < r < \infty, \end{cases} \tag{71}$$

*holds.*

(iii) *Let* $0 < p \leqslant 1$. *Then there exists a constant c such that*

$$\|\, I_{p,2}^m \,|\mathcal{L}_{r,\infty}^{(c)}\| \leqslant c \begin{cases} m^{1/r-1/2} & \text{if } 0 < r \leqslant 2, \\ 1 & \text{if } 2 < r < \infty, \end{cases} \tag{72}$$

*holds for all* $m \in \mathbb{N}$.

To prove the estimates of the Gelfand numbers from above, it turns out to be useful to split the identity $I$ into two parts $\mathrm{id}^1$, $\mathrm{id}^2$ and to treat them independently. In fact, we shall investigate $\|\,\mathrm{id}^i \,|\mathcal{L}_{r_i,\infty}^{(c)}\|$, $i = 1, 2$, where $r_1$ and $r_2$ are chosen in different ways. For basic properties of the Gelfand numbers we refer to Remark 7 and [10, 2.3].

**Theorem 8.** *Let* $0 < q \leqslant \infty$.

(i) *Let* $1 \leqslant p < 2$ *and suppose that* $t > d/2$. *Then*

$$d^n(I, b_{p,q}^{s+t}, b_{2,2}^s) \asymp n^{-t/d}.$$

(ii) *Let* $2 < p \leqslant \infty$ *and suppose that* $t > 0$. *Then*

$$d^n(I, b_{p,q}^{s+t}, b_{2,2}^s) \asymp n^{-t/d}.$$

(iii) *Let* $0 < p < 1$ *and suppose that*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right). \tag{73}$$

*Then there exist two constants* $c_1$ *and* $c_2$ *such that*

$$c_1 n^{-t/d} \leqslant d^n(I, b_{p,q}^{s+t}, b_{2,2}^s) \leqslant c_2 n^{-t/d-1+1/p}.$$

**Proof.** Without loss of generality we may assume $s = 0$. To see this consider the diagram

$$\begin{array}{ccc} b^{s+t}_{p,q} & \xrightarrow{I_1} & b^s_{2,2} \\ {\scriptstyle L_{-s}}\big\downarrow & & \big\uparrow{\scriptstyle L_s} \\ b^t_{p,q} & \xrightarrow{I_2} & b^0_{2,2}\,, \end{array}$$

where $L_s$ denotes the isomorphism introduced in Remark 17. The multiplicativity of the Gelfand numbers implies that

$$d^n(I_1, b^{s+t}_{p,q}, b^s_{2,2}) \leqslant \|L_{-s}\|\,\|L_s\| d^n(I_2, b^t_{p,q}, b^0_{2,2}),$$

comparing with Remark 7. Changing $L_{-s}$ into $L_s$ and vice versa in the diagram above we end up with

$$d^n(I_1, b^{s+t}_{p,q}, b^s_{2,2}) = d^n(I_2, b^t_{p,q}, b^0_{2,2}).$$

*Step* 1: Estimate from above. We concentrate on natural numbers $n = 2^{Nd}$ for $N \in \mathbb{N}$ (the remaining can be treated by the monotonicity of the $d^n$). Let $\mathrm{id}_j$ denote the projection given by

$$\big(\mathrm{id}_j\, a\big)_{m,\lambda} := \begin{cases} a_{j,\lambda} & \text{if } m = j, \\ 0 & \text{otherwise.} \end{cases}$$

We split the identity $I$ into a sum $I = \mathrm{id}^1 + \mathrm{id}^2$ depending on $N$, where

$$\mathrm{id}^1 := \sum_{j=0}^{N} \mathrm{id}_j \quad \text{and} \quad \mathrm{id}^2 := \sum_{j=N+1}^{\infty} \mathrm{id}_j.$$

Later on we shall apply the following observation. Consider the diagram

$$\begin{array}{ccc} b^t_{p,q}(\nabla) & \xrightarrow{\mathrm{id}_j} & b^0_{2,2}(\nabla) \\ {\scriptstyle P}\big\downarrow & & \big\uparrow{\scriptstyle Q} \\ \ell_p^{|\nabla_j|} & \xrightarrow{I^{|\nabla_j|}_{p,2}} & \ell_2^{|\nabla_j|}\,, \end{array}$$

where $P$ and $Q$ are defined as follows. Let $a = (a_{\ell,\lambda})_{\ell,\lambda}$. Then

$$(P(a))_\lambda := a_{j,\lambda}.$$

For $b = (b_\lambda)_\lambda$ we define

$$(Q(b))_{\ell,\lambda} := \begin{cases} a_{j,\lambda} & \text{if } j = \ell, \\ 0 & \text{otherwise.} \end{cases}$$

Obviously,

$$\|P\| = 2^{-j(t+d(1/2-1/p))} \quad \text{and} \quad \|Q\| = 1.$$

Then property (27) for the Gelfand numbers yields

$$
\begin{aligned}
d^n(\mathrm{id}_j, b^{s+t}_{p,q}, b^s_{2,2}) &\leqslant \|P\|\ \|Q\|\ d^n(I^{|\nabla_j|}_{p,2}) \\
&\leqslant 2^{-j(t+d(1/2-1/p))}\, d^n(I^{|\nabla_j|}_{p,2}).
\end{aligned} \tag{74}
$$

*Substep* 1.1. The estimate of $d^n(\mathrm{id}^1, b^t_{p,q}, b^0_{2,2})$, $n = 2^{Nd}$. First we suppose $2 \leqslant p \leqslant \infty$. Thanks to (69), (70), and (74) we find

$$
\begin{aligned}
\|\mathrm{id}^1 | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho &\leqslant \sum_{j=0}^{N} \|\,\mathrm{id}_j\, | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\
&\leqslant \sum_{j=0}^{N} 2^{-j(t+d(1/2-1/p))\varrho}\, \|I^{|\nabla_j|}_{p,2}\, | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\
&\leqslant c_1 \sum_{j=0}^{N} 2^{-j(t+d(1/2-1/p))\varrho}\, 2^{jd(1/r-1/p+1/2)\varrho} \\
&\leqslant c_2 2^{N(d/r-t)\varrho}
\end{aligned} \tag{75}
$$

if $d > tr$. Choosing $r$ small enough, we derive from the definition of $\mathcal{L}^{(c)}_{r,\infty}$ that

$$
d^n(\mathrm{id}^1) = d^{2^{Nd}}(\mathrm{id}^1) \leqslant c_3 2^{-Nt} = c_3 n^{-t/d}. \tag{76}
$$

Now we consider the case $1 \leqslant p < 2$. As above, but using (71) instead of (70), we find

$$
\|\mathrm{id}^1\, | \mathcal{L}^{(c)}_{r,\infty}\| \leqslant c_2 2^{N(d/r-t)}
$$

if $1/r > t/d$ and $1/r \geqslant 2$. Choosing $r$ small enough, we obtain

$$
d^{2^{Nd}}(\mathrm{id}^1) \leqslant c_4 2^{-Nt}. \tag{77}
$$

Finally, we investigate the case $0 < p < 1$. As above, we obtain

$$
d^{2^{Nd}}(\mathrm{id}^1) \leqslant c_5 2^{-N(t+d-d/p)} = c_5\, n^{-t/d-1+1/p}. \tag{78}
$$

*Substep* 1.2: The estimate of $d^n(\mathrm{id}^2, b^t_{p,q}, b^0_{2,2})$, where $n = 2^{Nd}$. Again we split our considerations into the three cases $p \geqslant 2$ and $1 \leqslant p < 2$ and $0 < p < 1$. First, let $2 \leqslant p \leqslant \infty$. Using (69), (70), and (74), we find that

$$
\begin{aligned}
\|\mathrm{id}^2\, | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho &\leqslant \sum_{j=N+1}^{\infty} \|\mathrm{id}_j\, | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\
&\leqslant \sum_{j=N+1}^{\infty} 2^{-j(t+d(1/2-1/p))\varrho}\, \|I^{|\nabla_j|}_{p,2}\, | \mathcal{L}^{(c)}_{r,\infty}\|^\varrho \\
&\leqslant c_1 \sum_{j=N+1}^{\infty} 2^{-j(t+d(1/2-1/p))\varrho}\, 2^{jd(1/r-1/p+1/2)\varrho} \\
&\leqslant c_2 2^{N(d/r-t)\varrho}
\end{aligned} \tag{79}
$$

if $tr > d$. Choosing $r$ large enough ($t > 0$ by assumption), we derive

$$d^{2^{Nd}}(\mathrm{id}^2) \leqslant c_3 2^{-Nt}. \tag{80}$$

Now we consider $1 \leqslant p < 2$. Similarly,

$$\| \mathrm{id}^2 \,|\mathcal{L}^{(c)}_{r,\infty}\| \leqslant c_3 2^{N(d/r-t)} \quad \text{if } \frac{1}{2} \leqslant \frac{1}{r} < \frac{t}{d}.$$

Since $t > d/2$, such a choice is always possible. Consequently,

$$d^{2^{Nd}}(\mathrm{id}^2) \leqslant c_4 2^{-Nt}. \tag{81}$$

Finally, let $0 < p < 1$. Then

$$d^{2^{Nd}}(\mathrm{id}^1) \leqslant c_5 2^{-N(t+d-d/p)} \quad \text{if } \frac{t}{d} + 1 - \frac{1}{p} > \frac{1}{r} \geqslant \frac{1}{2}. \tag{82}$$

Such a choice is always possible if (73) holds.

*Substep* 1.3: The additivity of the Gelfand widths yields

$$d^{2n}(\mathrm{id}) \leqslant d^n(\mathrm{id}^1) + d^n(\mathrm{id}^2).$$

In view of this inequality, the estimate from above of the Gelfand widths follows from (76)–(82).

*Step* 2: Estimate from below. Since $b_n \leqslant c d^n$, cf. Lemma 1(i), we may use Lemma 5 here to derive the lower bound in the case $0 < p \leqslant 2$. For $p > 2$, we shall use a different argument. Again we restrict ourselves to a subsequence of the natural numbers $n$, where

$$\frac{|\nabla_N|}{2} \leqslant n < \frac{|\nabla_N|}{2} + 1\,, \quad N \in \mathbb{N}.$$

Consider the diagram

$$\begin{array}{ccc} \ell_p^{|\nabla_N|} & \xrightarrow{I_1} & \ell_2^{|\nabla_N|} \\ P\big\downarrow & & \big\uparrow Q \\ b^t_{p,q}(\nabla) & \xrightarrow{I_2} & b^0_{2,2}(\nabla)\,, \end{array}$$

where $I_1$ and $I_2$ denote identities and this time $P$ and $Q$ are defined as follows. Let $b = (b_\lambda)_{\lambda\in\nabla_N}$. Then

$$(P(b))_{j,\lambda} := \begin{cases} b_\lambda & \text{if} \quad j = N, \\ 0 & \text{otherwise.} \end{cases}$$

For $a = (a_{j,\lambda})_{j,\lambda}$ we define

$$(Q(a))_\lambda := a_{N,\lambda}\,, \quad \lambda \in \nabla_N.$$

Obviously,

$$\| P \| = 2^{N(t+d(1/2-1/p))} \quad \text{and} \quad \|Q\| = 1.$$

Then property (27) for the Gelfand numbers yields that

$$d^n(I_1, \ell_p^{|\nabla_N|}, \ell_2^{|\nabla_N|}) \leqslant \|P\| \; \|Q\| \; d^n(I_2, b^t_{p,q}(\nabla), b^0_{2,2}(\nabla))$$

which, in view of Gluskin's estimates (67), implies that

$$c 2^{Nd(1/2-1/p)} \leqslant 2^{N(t+d(1/2-1/p))} d^n(I_2, b^t_{p,q}, b^0_{2,2})$$

for some positive $c$ (independent of $N$). This completes the estimate from below. □

**Remark 23.** The use of operator ideals in such a connection and the associated splitting technique applied in Step 1 has some history, cf. [9,56,58]. Closest to us is [56], where these methods have been used in connection with entropy numbers.

### 4.3. *Widths of embeddings of Besov spaces*

Here we do not formulate a general result, since the restrictions on the domains are different for different widths.

#### 4.3.1. *The manifold widths of the identity*

The main result of this subsection consists in the following nondiscrete counterpart of Theorem 6.

**Theorem 9.** *Let $\Omega$ be a bounded Lipschitz domain. Let $0 < p_0, p_1 \leqslant \infty$, $0 < q_0, q_1 \leqslant \infty$, and $s \in \mathbb{R}$. Suppose that* (55) *holds. Then we have*

$$e_n^{\text{cont}}(I, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega))) \asymp n^{-t/d} \,. \tag{83}$$

**Remark 24.** Theorem 9 has several forerunners. We would like to mention [30,32,40]. In these papers, the authors consider the quantities $e_n^{\text{cont}}(I, B^t_{q_0}(L_{p_0}(\Omega)), L_{p_1}(\Omega))$. Note that from the continuous embeddings

$$B^0_1(L_p(\Omega)) \hookrightarrow L_p(\Omega) \hookrightarrow B^0_\infty(L_p(\Omega)), \quad 1 \leqslant p \leqslant \infty,$$

we obtain as a direct consequence of Theorem 9

$$e_n^{\text{cont}}(I, B^t_{q_0}(L_{p_0}(\Omega)), L_{p_1}(\Omega)) \asymp n^{-t/d} \,, \tag{84}$$

as long as $1 \leqslant p_1 \leqslant \infty$ and $t > (1/p_0 - 1/p_1)_+$. So, Theorem 9 covers the results obtained before. However, let us mention that we used the ideas from [32] for our estimate from above and the ideas from [40] to derive the estimate from below (here on the level of sequence spaces).

**Proof of Theorem 9.** Let $\mathcal{E}$ denote a universal bounded linear extension operator corresponding to $\Omega$, see Proposition 6 in Appendix A.5. Let diam $\Omega$, be the diameter of $\Omega$ and let $x^0$ be a point in $\mathbb{R}^d$ such that

$$\Omega \subset \{y : \; |x^0 - y| \leqslant \operatorname{diam} \Omega\}.$$

Without loss of generality, we assume that

$$\operatorname{supp} \mathcal{E} f \subset \{y : \; |x^0 - y| \leqslant 2 \operatorname{diam} \Omega\}.$$

Let $\nabla$ be defined as in (99) and (100) (with $\Omega$ replaced by the ball with radius 2 diam $\Omega$ and center $x^0$). Let $R$ denote the restriction operator with respect to $\Omega$. Let $T$ denote the continuous linear operator that associates to $f$ its wavelet series; $T^{-1}$ is the inverse operator. Here we assume that we can characterize the Besov spaces $B^{s+t}_{p_0,q_0}(\mathbb{R}^d)$, as well as $B^s_{p_1,q_1}(\mathbb{R}^d)$, in the sense of Proposition 5 in Appendix A.3. Then we consider the diagram

$$\begin{array}{ccccc} B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{\mathcal{E}} & B^{s+t}_{q_0}(L_{p_0}(\mathbb{R}^d)) & \xrightarrow{T} & b^{s+t}_{p_0,q_0}(\nabla) \\ I_1 \quad \downarrow & & & & \downarrow \quad I_2 \\ B^s_{q_1}(L_{p_1}(\Omega)) & \xleftarrow{R} & B^s_{q_1}(L_{p_1}(\mathbb{R}^d)) & \xleftarrow{T^{-1}} & b^s_{p_1,q_1}(\nabla)\,. \end{array} \tag{85}$$

Observe that $I_1 = R \circ T^{-1} \circ I_2 \circ T \circ \mathcal{E}$. From (85) and (27) for $e^{\text{cont}}$, we derive that

$$e_n^{\text{cont}}(I_1, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega))) \leqslant \|\mathcal{E}\|\,\|T\|\,\|T^{-1}\|\, e_n^{\text{cont}}(I_2, b^{s+t}_{p_0,q_0}(\nabla), b^s_{p_1,q_1}(\nabla)).$$

For the converse inequality, we choose $\nabla^* = (\nabla^*_j)_j$ such that

$$\operatorname{supp} \psi_{j,\lambda} \subset \Omega, \quad \lambda \in \nabla^*_j, \quad j = -1, 0, 1, \ldots,$$

and $\inf_j \, 2^{-jd}\, |\nabla^*_j| > 0$. Then we consider the diagram

$$\begin{array}{ccc} b^{s+t}_{p_0,q_0}(\nabla^*) & \xrightarrow{\quad I_2 \quad} & b^s_{p_1,q_1}(\nabla^*) \\ T^{-1}\big\downarrow & & \big\uparrow T \\ B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{\quad I_1 \quad} & B^s_{q_1}(L_{p_1}(\Omega)), \end{array} \tag{86}$$

and conclude that

$$e_n^{\text{cont}}(I_2, b^{s+t}_{p_0,q_0}(\nabla^*), b^s_{p_1,q_1}(\nabla^*)) \leqslant \|T\|\,\|T^{-1}\|\, e_n^{\text{cont}}(I_1, B^{s+t}_{q_0}(L_{p_0}(\Omega)), B^s_{q_1}(L_{p_1}(\Omega))).$$

Now Theorem 6 yields the desired result. $\square$

*4.3.2. The widths of best n-term approximation of the identity*

Let $\Omega$ be a bounded Lipschitz domain in $\mathbb{R}^d$. We assume that for any fixed triple $(t, p, q)$ of parameters the spaces $B^{s+t}_q(L_p(\Omega))$ and $B^s_2(L_2(\Omega))$ allow a discretization by one common wavelet system $\mathcal{B}^*$. More exactly, we assume that (107)–(112) are satisfied simultaneously for both spaces, cf. Appendix A.10. From this, it follows that $\mathcal{B}^* \in \mathcal{B}_{C^*}$ for some $1 \leqslant C^* < \infty$.

**Theorem 10.** *Let $\Omega$ be as above. Let $0 < p \leqslant \infty$, $0 < q \leqslant \infty$, $s \in \mathbb{R}$ and*

$$t > d\left(\frac{1}{p} - \frac{1}{2}\right)_+$$

*holds. Then, for any $C \geqslant C^*$ we have*

$$e^{\text{non}}_{n,C}(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-t/d}.$$

**Remark 25.** (i) Periodic versions on the $d$-dimensional torus $T^d$ may be found in [82,83] with $B^s_2(L_2(\Omega))$ replaced by $L_{p_1}(T^d)$ and $p_1, p, q \geqslant 1$. Furthermore, more general classes of functions are investigated there (anisotropic Besov spaces, functions of dominating mixed smoothness).

Finally, let us mention that estimates from below for the quantities

$$\inf_{\mathcal{B}\in\mathcal{O}} \sup_{\|u\|_{B^t_{q_1}(L_{p_1}(T^d))}\leqslant 1} \sigma_n(u,\mathcal{B})_{L_2(T^d)},$$

where $\mathcal{O}$ is the set of all orthonormal bases, have been given by Kashin ($p_1 = q_1 = \infty,\ d = 1$) and Temlyakov [82,83] (general anisotropic case). Instead of the manifold widths these authors use entropy numbers.

(ii) We stress that, in this paper, we study the approximation in some Hilbertian smoothness space $B^s_2(L_2(\Omega))$ while most known results from the literature concern approximation in an $L_p(\Omega)$-space.

**Remark 26.** We also recall the following limiting case. Let $0 < p < 2$ and $t = d(1/p - 1/2)$. Then the embedding $B^{s+t}_p(L_p(\Omega)) \hookrightarrow B^s_2(L_2(\Omega))$ is continuous but not compact, cf. Proposition 7. Here we have

$$\left(\sum_{n=1}^{\infty}\left[n^{t/d}\,\sigma_n(u,\mathcal{B}^*)_{B^s_2(L_2(\Omega))}\right]^p \frac{1}{n}\right)^{1/p} < \infty \quad \text{if and only if } u \in B^{s+t}_p(L_p(\Omega)).$$

A proof can be found in [20, Proposition 1], but the argument there is mainly based on [34], see also [31].

**Proof of Theorem 10.** Let $\mathcal{B}^*$ be a wavelet basis as in Appendix A.10. Let $\mathcal{B}$ denote the canonical orthonormal basis of $b^0_{2,2}(\nabla)$. We equip the Besov space with the equivalent quasi-norm (112). Observe,

$$\sigma_n(f,\mathcal{B}^*)_{B^s_{p_1,q_1}(\Omega)} \leqslant c\sigma_n((\langle f,\widetilde{\psi}_{j,\lambda}\rangle)_{j,\lambda},\mathcal{B})_{b^s_{p_1,q_1}(\nabla)},$$

where $c$ is one of the constants in (111). By means of Theorem 6 and Remark 2(iii), this implies the estimate from above. The estimate from below follows by combining Theorems 1 and 9. $\square$

The simple arguments used in the proof of Theorem 10 allow us to carry over Remark 26 to the sequence space level, see Remark 18, and Theorem 7 to the level of function spaces.

**Theorem 11.** *Let $\Omega$ and $\mathcal{B}^*$ be as above. Let $0 < p_0, p_1, q_0, q_1 \leqslant \infty$, $s \in \mathbb{R}$ and $t > 0$ such that* (55) *holds. Then we have*

$$\sup\left\{\sigma_n(u,\mathcal{B}^*)_{B^s_{q_1}(L_{p_1}(\Omega))} \ : \quad \|\, u|B^{s+t}_{q_0}(L_{p_0}(\Omega))\| \leqslant 1\right\} \asymp n^{-t/d}.$$

**Remark 27.** (i) For earlier results in this direction we refer to [33,38,54,68].

(ii) Not all orthonormal systems are of the same quality, see [38]. Let us mention the following result of DeVore and Temlyakov [36]. Let $\mathcal{B}^\#$ denote the trigonometric system in $\mathbb{R}^d$. By $B^s_q(L_p(\mathbb{T}^d))$ we mean the periodic Besov spaces defined on the $d$-dimensional torus $\mathbb{T}^d$. Then we put

$$t(p_0,p_1) := \begin{cases} d\big(1/p_0 - 1/p_1\big)_+ & \text{if } 0 < p_0 \leqslant p_1 \leqslant 2 \text{ or } 1 \leqslant p_1 \leqslant p_0 \leqslant \infty, \\ d\max\,(1/p_0, 1/2) & \text{otherwise.} \end{cases}$$

If $1 \leqslant p_1 \leqslant \infty$, $0 < p_0, q_0 \leqslant \infty$, and $t > t(p_0, p_1)$, then

$$\sup \left\{ \sigma_n(u, \mathcal{B}^{\#})_{L_{p_1}(\mathbb{T}^d)} : \| u \,|B^t_{q_0}(L_{p_0}(\mathbb{T}^d))\| \leqslant 1 \right\} \asymp \begin{cases} n^{-t/d} & \text{if } p_0 \geqslant \max(p_1, 2), \\ n^{-t/d+1/p_0-1/2} & \text{if } p_0 \leqslant \max(p_1, 2) = 2, \\ n^{-t/d+1/p_0-1/p_1} & \text{if } p_0 \leqslant \max(p_1, 2) = p_1. \end{cases}$$

*4.3.3. The approximation numbers of the identity*

**Theorem 12.** *Let $\Omega$ be a bounded Lipschitz domain. Let $0 < p \leqslant \infty$, $0 < q \leqslant \infty$, and $s \in \mathbb{R}$.* Suppose that

$$t > d \left( \frac{1}{p} - \frac{1}{2} \right)_+$$

*holds. Then we have*

$$e^{\text{lin}}_n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp \begin{cases} n^{-t/d} & \textit{if } 2 \leqslant p \leqslant \infty, \\ n^{-t/d+1/p-1/2} & \textit{if } 0 < p < 2. \end{cases}$$

**Proof.** The statement is a consequence of Theorem 6(ii), Proposition 6, (101) and (102). □

**Remark 28.** (i) The proof is constructive. An order-optimal linear approximation is obtained by taking an appropriate partial sum of the wavelet series of $\mathcal{E} f$, where $\mathcal{E}$ is the linear universal extension operator from Proposition 6, cf. Remark 22 for the discrete case.

(ii) This result is well known. It can be derived from [91,43, 3.3.2]. There and in [7] information can also be found about what is known for the general situation, i.e., in which $B^s_2(L_2(\Omega))$ is replaced by $B^s_{q_1}(L_{p_1}(\Omega))$. However, let us mention that there are many references which had dealt with this problem before; we refer to [81, Theorem. 1.4.2, 85, Theorem 9, p. 193] and the comments given there.

*4.3.4. The Gelfand widths of the identity*

**Theorem 13.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain and let $0 < q \leqslant \infty$.*

(i) *Let $1 \leqslant p < 2$ and suppose that $t > d/2$. Then*

$$d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-t/d}.$$

(ii) *Let $2 < p \leqslant \infty$ and suppose that $t > 0$. Then*

$$d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \asymp n^{-t/d}.$$

(iii) *Let $0 < p < 1$ and suppose that*

$$t > d \left( \frac{1}{p} - \frac{1}{2} \right).$$

*Then there exists two constants $c_1$ and $c_2$ such that*

$$c_1 n^{-t/d} \leqslant d^n(I, B^{s+t}_q(L_p(\Omega)), B^s_2(L_2(\Omega))) \leqslant c_2 n^{-t/d-1+1/p}.$$

**Proof.** Consider the diagram

$$\begin{array}{ccc} B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{I_1} & B^s_2(L_2(\Omega)) \\ T\big\downarrow & & \big\uparrow T^{-1} \\ b^{s+t}_{p_0,q_0}(\nabla) & \xrightarrow{I_2} & b^s_{2,2}(\nabla)\,, \end{array}$$

where $T$ and $T^{-1}$ are defined as in the proof of Theorem 9. Since $I_1 = T^{-1} \circ I_2 \circ T$, it is enough to combine property (27) for the Gelfand numbers and Theorem 8 to derive the estimates from above. For the estimates from below, one uses the diagram

$$\begin{array}{ccc} b^{s+t}_{p_0,q_0}(\nabla^*) & \xrightarrow{I_1} & b^s_{2,2}(\nabla^*) \\ T\big\downarrow & & \big\uparrow T^{-1} \\ B^{s+t}_{q_0}(L_{p_0}(\Omega)) & \xrightarrow{I_2} & B^s_2(L_2(\Omega))\,, \end{array}$$

where $\nabla^*$ is defined as in proof of Theorem 9. This completes the proof. $\square$

**Remark 29.** Partial results concerning Gelfand numbers of embedding operators may be found in the monographs [73, Chapter VII, Theorem 1.1, 85, Theorem 39, p. 206, 88, 4.10.2]. Let $T$ be a compact operator in $\mathcal{L}(F, E)$, where $F, E$ are arbitrary Banach spaces and let $d_n(T, F, E)$ denote the Kolmogorov numbers. Then

$$d^n(T') = d_n(T)\,, \quad n \in \mathbb{N},$$

holds, cf. [10, Proposition 2.5.6] or [71]. For Kolmogorov numbers the asymptotic behavior is also known in certain situations, cf. [73, Chapter VII, Theorem 1.1, 85, Theorem 10, p. 193, 88, 4.10.2, 81].

### 4.4. *Proofs of Theorems 2, 4, and 5*

#### 4.4.1. *Proof of Theorem 2*

For $s > 0$ we have $H^{-s}(\Omega) = B^{-s}_2(L_2(\Omega))$. Hence, Theorem 12 yields

$$e^{\text{lin}}_n(I, B^{-s+t}_q(L_p(\Omega)), H^{-s}(\Omega)) \asymp \begin{cases} n^{-t/d} & \text{if } 0 < p \leqslant 2, \\ n^{-t/d+1/p-1/2} & \text{if } 2 < p \leqslant \infty. \end{cases}$$

Since $S : H^{-s}(\Omega) \to H^s_0(\Omega)$ is an isomorphism, we obtain the desired result from property (27) for the approximation numbers.

#### 4.4.2. *Proof of Theorem 4*

Since $H^{-s}(\Omega) = B^{-s}_2(L_2(\Omega))$, Theorem 10 yields that

$$e^{\text{non}}_{n,C}(I, B^{-s+t}_q(L_p(\Omega)), H^{-s}(\Omega)) \asymp n^{-t/d}.$$

Since $S : H^{-s}(\Omega) \to H^s_0(\Omega)$ is an isomorphism, Lemma 3(ii) implies the desired result.

*4.4.3. Proof of Theorem 5*

All what we need from the wavelet basis is the following estimate for the best $n$-term approximation in the $H^1$-norm:

$$\|u - S_n(f)\|_{H^1(\Omega)} \leqslant c\|u\,|B_\tau^{t+1}(L_\tau(\Omega))\|\, n^{-t/2}, \quad \text{where } \frac{1}{\tau} = \frac{t}{2} + \frac{1}{2}, \tag{87}$$

see, e.g., [20] (however we could instead use Theorem 11). We therefore have to estimate the Besov norm $B_\tau^\alpha(L_\tau(\Omega))$. Since $1 < p \leqslant 2$, the embedding $B_p^{k-1}(L_p(\Omega)) \hookrightarrow W_p^{k-1}(\Omega)$ holds, cf. e.g., [89, 2.3.2, 2.5.6]. Hence our right-hand side $f$ is contained in the Sobolev space $W_p^{k-1}(\Omega)$. Therefore we may employ the fact that $u$ can be decomposed into a regular part $u_{\mathrm{R}}$ and a singular part $u_{\mathrm{S}}$, i.e., $u = u_{\mathrm{R}} + u_{\mathrm{S}}$, where $u_{\mathrm{R}} \in W_p^{k+1}(\Omega)$ and $u_{\mathrm{S}}$ only depends on the shape of the domain and can be computed explicitly, cf. [49, Theorem. 2.4.3]. We introduce polar coordinates $(r_l, \theta_l)$ in the vicinity of each vertex $\Upsilon_l$ and introduce the functions

$$\mathcal{S}_{l,m}(r_l,\theta_l) := \begin{cases} \zeta_l(r_l) r_l^{\lambda_{l,m}} \sin(m\pi\theta_l/\omega_l) \\ \quad \text{if } \lambda_{l.m} := m\pi/\omega_l \neq \text{integer}, \\ \\ \zeta_l(r_l) r_l^{\lambda_{l,m}} [\log r_l \sin(m\pi\theta_l/\omega_l) + \theta_l \cos(m\pi\theta_l/\omega_l)] \\ \quad \text{otherwise.} \end{cases}$$

Here $\zeta_1, \ldots, \zeta_N$ denote suitable $C^\infty$ truncation functions and $m$ is a natural number. Then for $f \in W_p^{k-1}(\Omega)$, one has

$$u_{\mathrm{S}} = \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} c_{l,m}\, \mathcal{S}_{l,m}, \tag{88}$$

provided that no $\lambda_{l,m}$ is equal to $k + 1 - 2/p$. This means that the finite number of singularity functions that is needed depends on the scale of spaces we are interested in, i.e., on the smoothness parameter $k$. According to (87), we have to estimate the Besov regularity of both, $u_{\mathrm{S}}$ and $u_{\mathrm{R}}$, in the specific scale

$$B_\tau^{t+1}(L_\tau(\Omega)) \quad \text{where } \frac{1}{\tau} = \frac{t}{2} + \frac{1}{2}.$$

Since $u_{\mathrm{R}} \in W_p^{k+1}(\Omega)$, the boundedness of $\Omega$ implies the embedding

$$W_p^{k+1}(\Omega) \hookrightarrow B_q^{k+1-\delta}(L_q(\Omega)) \quad \text{with } \delta > 0,\ 0 < q \leqslant p,\ k+1 > 2\left(\frac{1}{q} - \frac{1}{2}\right).$$

Hence

$$u_{\mathrm{R}} \in B_\tau^{k+1-\delta}(L_\tau(\Omega)) \quad \text{with } \frac{1}{\tau} = \frac{(k-\delta)}{2} + \frac{1}{2} \text{ for arbitrarily small } \delta > 0. \tag{89}$$

Moreover, it has been shown in [17] (see also Remark 31) that the functions $\mathcal{S}_{l,m}$ defined above satisfy

$$\mathcal{S}_{l,m}(r_l,\theta_l) \in B_q^{1/2+2/q}(L_q(\Omega)) \quad \text{for all } 0 < q < \infty. \tag{90}$$

By combining (89) and (90) we see that

$$u \in B^{k+1-\delta}_{\tau}(L_\tau(\Omega)) \quad \text{where } \frac{1}{\tau} = \frac{(k-\delta)}{2} + \frac{1}{2} \text{ for arbitrarily small } \delta > 0.$$

To derive an estimate uniformly with respect to the unit ball in $B^{k-1}_p(L_p(\Omega))$ we argue as follows. We put

$$\mathcal{N} := \operatorname{span}\left\{\mathcal{S}_{l,m}(r_l, \theta_l) : \quad 0 < \lambda_{m,l} < k+1-2/p\,,\ l = 1, \ldots, N\right\}.$$

Let $\gamma_l$ be the trace operator with respect to the segment $\Gamma_l$. Grisvard has shown that $\Delta$ maps

$$H := \left\{u \in W^{k+1}_p(\Omega) : \quad \gamma_l u = 0\,,\ l = 1, \ldots, N\right\} + \mathcal{N}$$

onto $W^{k-1}_p(\Omega)$, cf. [48, Theorem. 5.1.3.5]. This mapping is also injective, see [48, Lemma 4.4.3.1, Remark. 5.1.3.6]. We equip the space $H$ with the norm

$$\|u\|_H := \|u_{\mathrm{R}} + u_{\mathrm{S}}\|_H = \|u_{\mathrm{R}}\|_{W^{k+1}_p(\Omega)} + \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} |c_{l,m}|,$$

see (88). Then $H$ becomes a Banach space. Furthermore, $\Delta : H \to W^{k-1}_p(\Omega)$ is continuous. Banach's continuous inverse theorem implies that the solution operator is continuous, considered as a mapping from $W^{k-1}_p(\Omega)$ onto $H$. Finally, observe that

$$\|u_{\mathrm{R}} + u_{\mathrm{S}}\|_{B^{k+1-\delta}_{\tau}(L_\tau(\Omega))} \leqslant c \left( \| u_{\mathrm{R}} \|_{W^{k+1}_p(\Omega)} + \sum_{l=1}^{N} \sum_{0<\lambda_{l,m}<k+1-2/p} |c_{l,m}| \right)$$

with some constant $c$ independent of $u$. $\square$

## Appendix A. Besov spaces

Here we collect some properties of Besov spaces that have been used in the text before. Detailed references will be given. For general information on Besov spaces, we refer to the monographs [62,63,69,74,89,90].

### A.1. *Besov spaces on $\mathbb{R}^d$ and differences*

Nowadays Besov spaces are widely used in several branches of mathematics. Probably the most common way to introduce these classes makes use of differences. For $M \in \mathbb{N}$, $h \in \mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{C}$ we define

$$\Delta^M_h f(x) := \sum_{j=0}^{M} \binom{M}{j} (-1)^{M-j} f(x + jh).$$

Let $0 < p \leqslant \infty$. The corresponding modulus of smoothness is then given by

$$\omega^M(t, f)_p := \sup_{|h|<t} \|\Delta^M_h f\|_{L_p(\mathbb{R}^d)}, \quad t > 0.$$

One approach to introduce Besov spaces is the following.

**Definition 4.** Let $s > 0$ and $0 < p, q \leqslant \infty$. Let $M$ be a natural number satisfying $M > s$. Then $\Lambda^s_q(L_p(\mathbb{R}^d))$ is the collection of all functions $f \in L_p(\mathbb{R}^d)$ such that

$$|f|_{\Lambda^s_q(L_p(\mathbb{R}^d))} := \left( \int_0^\infty \left[ t^{-s}\, \omega^M(t, f)_p \right]^q \frac{dt}{t} \right)^{1/q} < \infty$$

if $q < \infty$ and

$$|f|_{\Lambda^s_\infty(L_p(\mathbb{R}^d))} := \sup_{t>0} t^{-s}\, \omega^M(t, f)_p < \infty$$

if $q = \infty$. These classes are equipped with a quasi-norm by taking

$$\|f\|_{\Lambda^s_q(L_p(\mathbb{R}^d))} := \|f\|_{L_p(\mathbb{R}^d)} + |f|_{\Lambda^s_q(L_p(\mathbb{R}^d))}.$$

**Remark 30.** It turns out that these classes do not depend on *M*, cf. [35].

**Remark 31.** Let $\varrho \in C_0^\infty(\mathbb{R}^d)$ be a function such that $\varrho(0) \neq 0$. By means of the above definition it is not complicated to show that a function

$$f_\alpha(x) := |x|^\alpha\, \varrho(x)\,, \quad x \in \mathbb{R}^d\,, \;\; \alpha > 0,$$

belongs to $\Lambda^{\alpha+d/p}_\infty(L_p(\mathbb{R}^d))$ and that this is the best possible (if $\alpha$ is not an even natural number), cf. [74, 2.3.1] for details. A minor modification shows that

$$f_{\alpha,\beta}(x) := |x|^\alpha\, (\log|x|)^\beta\, \varrho(x)\,, \qquad x \in \mathbb{R}^d\,, \;\; \alpha,\; \beta > 0,$$

belongs to $\Lambda^{\alpha+d/p-\varepsilon}_\infty(L_p(\mathbb{R}^d))$ for all $\varepsilon$, $0 < \varepsilon < \alpha + d/p$.

### *A.2. Besov spaces on $\mathbb{R}^d$ and Littlewood–Paley characterizations*

Since we are using also spaces with negative smoothness $s < 0$ and/or $p, q < 1$ we shall give a further definition, which relies on Fourier analysis. We use it here for introductory purposes. This approach makes use of smooth dyadic decompositions of unity. Let $\varphi \in C_0^\infty(\mathbb{R}^d)$ be a function such that $\varphi(x) = 1$ if $|x| \leqslant 1$ and $\varphi(x) = 0$ if $|x| \geqslant 2$. Then we put

$$\varphi_0(x) := \varphi(x),\;\; \varphi_j(x) := \varphi(2^{-j}x) - \varphi(2^{-j+1}x), \quad j \in \mathbb{N}. \tag{91}$$

It follows

$$\sum_{j=0}^\infty \varphi_j(x) = 1\,, \quad x \in \mathbb{R}^d$$

and

$$\operatorname{supp} \varphi_j \subset \left\{ x \in \mathbb{R}^d : \quad 2^{j-2} \leqslant |x| \leqslant 2^{j+1} \right\}, \quad j = 1, 2, \ldots\,.$$

Let $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, both defined on $\mathcal{S}'(\mathbb{R}^d)$. For $f \in \mathcal{S}'(\mathbb{R}^d)$ we consider the sequence $\mathcal{F}^{-1}[\varphi_j(\xi)\, \mathcal{F}f(\xi)](x)$, $j \in \mathbb{N}_0$, of entire analytic functions. By means of these functions, we define the Besov classes.

**Definition 5.** Let $s \in \mathbb{R}$ and $0 < p, q \leqslant \infty$. Then $B^s_q(L_p(\mathbb{R}^d))$ is the collection of all tempered distributions $f$ such that

$$\| f \,|B^s_q(L_p(\mathbb{R}^d))\| = \left( \sum_{j=0}^{\infty} 2^{sjq} \, \| \mathcal{F}^{-1}[\varphi_j(\xi)\, \mathcal{F} f(\xi)](\,\cdot\,) \,| L_p(\mathbb{R}^d) \|^q \right)^{1/q} < \infty$$

if $q < \infty$ and

$$\| f | B^s_\infty(L_p(\mathbb{R}^d))\| = \sup_{j=0,1,\ldots} \; 2^{sj} \, \| \mathcal{F}^{-1}[\varphi_j(\xi)\, \mathcal{F} f(\xi)](\,\cdot\,) \,| L_p(\mathbb{R}^d) \| < \infty$$

if $q = \infty$.

**Remark 32.** (i) If no confusion is possible we drop $\mathbb{R}^d$ in notations.

(ii) These classes are quasi-Banach spaces. They do not depend on the chosen function $\varphi$ (up to equivalent quasi-norms). If $t = \min(1, p, q)$, then

$$\| f + g \,|B^s_q(L_p)\|^t \leqslant \| f \,|B^s_q(L_p)\|^t + \|g \,|B^s_q(L_p)\|^t$$

holds for all $f, g \in B^s_q(L_p)$.

**Proposition 4** (*Triebel [89, 2.5.12]*). Let $0 < p, q \leqslant \infty$ and $s > d \max(0, 1/p - 1)$. Then we have coincidence of $\Lambda^s_q(L_p)$ and $B^s_q(L_p)$ in the sense of equivalent quasi-norms.

**Remark 33.** (i) For $s \leqslant d \max(0, 1/p - 1)$ we have $\Lambda^s_q(L_p) \neq B^s_q(L_p)$. E.g., the Dirac distribution $\delta$ belongs to $B^{d(1/p-1)}_\infty(L_p)$, cf. [74, 2.3.1].

(ii) Smooth cut-off functions are pointwise multipliers for all Besov spaces. More exactly, let $\psi \in \mathcal{D}$. Then the product $\psi f$ belongs to $B^s_q(L_p)$ for any $f \in B^s_q(L_p)$ and there exists a constant $c$ such that

$$\|\psi f \,|B^s_q(L_p)\| \leqslant c \| f \,|B^s_q(L_p)\|$$

holds, see e.g., [89, 2.8, 74, 4.7].

### *A.3. Wavelet characterizations*

For the construction of biorthogonal wavelet bases as considered below, we refer to the recent monograph of Cohen [12, Chapter 2]. Let $\varphi$ be a compactly supported scaling function of sufficiently high regularity and let $\psi_i$, where $i = 1, \ldots, 2^d - 1$, be the corresponding wavelets. More exactly, we suppose for some $N > 0$ and $r \in \mathbb{N}$

$$\begin{aligned}
&\operatorname{supp}\varphi\,, \operatorname{supp}\psi_i \quad \subset \quad [-N, N]^d\,, \quad i = 1, \ldots, 2^d - 1\,,\\
&\varphi, \psi_i \in C^r(\mathbb{R}^d)\,, \quad i = 1, \ldots, 2^d - 1,\\
&\int x^\alpha \, \psi_i(x)\, dx = 0 \quad \text{for all } |\alpha| \leqslant r\,, \quad i = 1, \ldots, 2^d - 1,
\end{aligned}$$

and

$$\varphi(x - k),\; 2^{jd/2}\, \psi_i(2^j x - k)\,, \quad j \in \mathbb{N}_0\,, \;\; k \in \mathbb{Z}^d,$$

is a Riesz basis in $L_2(\mathbb{R}^d)$. We shall use the standard abbreviations

$$\psi_{i,j,k}(x) = 2^{jd/2}\,\psi_i(2^j x - k) \quad \text{and} \quad \varphi_k(x) = \varphi(x-k)\,.$$

Further, the dual Riesz basis should fulfill the same requirements, i.e., there exist functions $\widetilde{\varphi}$ and $\widetilde{\psi}_i$, $i = 1, \ldots, 2^d - 1$, such that

$$\begin{aligned}
&\langle \widetilde{\varphi}_k, \psi_{i,j,k}\rangle = \langle \widetilde{\psi}_{i,j,k}, \varphi_k\rangle = 0,\\
&\langle \widetilde{\varphi}_k, \varphi_\ell\rangle = \delta_{k,\ell} \quad \text{(Kronecker symbol)}\,,\\
&\langle \widetilde{\psi}_{i,j,k}, \psi_{u,v,\ell}\rangle = \delta_{i,u}\,\delta_{j,v}\,\delta_{k,\ell},\\
&\operatorname{supp}\widetilde{\varphi}\,, \operatorname{supp}\widetilde{\psi}_i \quad \subset \quad [-N, N]^d, \quad i = 1, \ldots, 2^d - 1,\\
&\widetilde{\varphi}, \widetilde{\psi}_i \in C^r(\mathbb{R}^d), \qquad i = 1, \ldots, 2^d - 1,\\
&\int x^\alpha\, \widetilde{\psi}_i(x)\, dx = 0 \quad \text{for all } |\alpha| \leqslant r,\ i = 1, \ldots, 2^d - 1.
\end{aligned}$$

For $f \in \mathcal{S}'(\mathbb{R}^d)$ we put

$$\langle f, \psi_{i,j,k}\rangle = f(\overline{\psi_{i,j,k}}) \quad \text{and} \quad \langle f, \varphi_k\rangle = f(\overline{\varphi_k}), \tag{92}$$

whenever this makes sense.

**Proposition 5.** *Let $s \in \mathbb{R}$ and $0 < p, q \leqslant \infty$. Suppose*

$$r > \max\left(s, \frac{2d}{p} + \frac{d}{2} - s\right). \tag{93}$$

*Then $B^s_q(L_p)$ is the collection of all tempered distributions f such that f is representable as*

$$f = \sum_{k\in\mathbb{Z}^d} a_k\,\varphi_k + \sum_{i=1}^{2^d-1}\sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}^d} a_{i,j,k}\,\psi_{i,j,k} \quad (\textit{convergence in} \quad \mathcal{S}')$$

*with*

$$\|f\,|B^s_q(L_p)\|^* := \left(\sum_{k\in\mathbb{Z}^d} |a_k|^p\right)^{1/p} + \left(\sum_{i=1}^{2^d-1}\sum_{j=0}^{\infty} 2^{j(s+d(1/2-1/p))q}\left(\sum_{k\in\mathbb{Z}^d}|a_{i,j,k}|^p\right)^{q/p}\right)^{1/q} < \infty$$

*if $q < \infty$ and*

$$\|f\,|B^s_\infty(L_p)\|^* := \left(\sum_{k\in\mathbb{Z}^d} |a_k|^p\right)^{1/p} + \sup_{i=1,\ldots,2^d-1}\ \sup_{j=0,\ldots}\ 2^{j(s+d(1/2-1/p))}\left(\sum_{k\in\mathbb{Z}^d}|a_{i,j,k}|^p\right)^{1/p} < \infty\,.$$

*The representation is unique and*

$$a_{i,j,k} = \langle f, \widetilde{\psi}_{i,j,k}\rangle \quad \text{and} \quad a_k = \langle f, \widetilde{\varphi}_k\rangle$$

*hold. Further* $I : f \mapsto \{\langle f, \widetilde{\varphi}_k\rangle, \langle f, \widetilde{\psi}_{i,j,k}\rangle\}$ *is an isomorphic map of* $B^s_q(L_p(\mathbb{R}^d))$ *onto the sequence space equipped with the quasi-norm* $\| \cdot |B^s_q(L_p)\|^*$, *i.e.,* $\| \cdot |B^s_q(L_p)\|^*$ *may serve as an equivalent quasi-norm on* $B^s_q(L_p)$.

**Remark 34.** (i) The restriction (93) guarantees that (92) makes sense for all $f \in B^s_q(L_p)$.

(ii) It is immediate from this proposition that the functions $\varphi_k, \psi_{i,j,k}, k \in \mathbb{Z}^d, 1 \leqslant i \leqslant 2^d-1, j \in \mathbb{N}_0$, form a basis for $B^s_q(L_p)$ if $\max(p,q) < \infty$. By the same reasoning the functions

$$\varphi_k, \quad 2^{-js}\,\psi_{i,j,k}, \quad k \in \mathbb{Z}^d,\ 1 \leqslant i \leqslant 2^d - 1,\ j \in \mathbb{N}_0\,,$$

form a Riesz basis for $B^s_2(L_2)$.

(iii) If the wavelet basis is orthonormal (in $L_2$), then this proposition is proved in [92]. But the comments made in Section 3.4 of the quoted paper make clear that this extends to the situation considered in Proposition 5. A different proof, but restricted to $s > d(1/p-1)_+$, is given in [12, Theorem 3.7.7]. However, there are many forerunners with some restrictions concerning $s$, $p$ and $q$. We refer to [6,62].

### *A.4. Besov spaces on domains—the approach via restrictions*

There are at least two different approaches to define function spaces on domains. One approach uses restrictions to $\Omega$ of functions defined on $\mathbb{R}^d$. So, all calculations are done on $\mathbb{R}^d$. The other approach introduces these spaces by means of local quantities defined only in $\Omega$. For numerical purposes the second approach is more promising whereas for analytic investigations the first one looks more elegant. Here we discuss both, since both were used.

Let $\Omega \subset \mathbb{R}^d$ be a bounded open nonempty set. Then we define $B^s_q(L_p(\Omega))$ to be the collection of all distributions $f \in \mathcal{D}'(\Omega)$ such that there exists a tempered distribution $g \in B^s_q(L_p(\mathbb{R}^d))$ satisfying

$$f(\varphi) = g(\varphi) \quad \text{for all} \;\; \varphi \in \mathcal{D}(\Omega),$$

i.e., $g|_\Omega = f$ in $\mathcal{D}'(\Omega)$. We put

$$\|f\,|B^s_q(L_p(\Omega))\| := \inf \|\,g\,|B^s_q(L_p(\mathbb{R}^d))\|,$$

where the infimum is taken with respect to all distributions $g$ as above.
Let $\operatorname{diam}\Omega$ be the diameter of the set $\Omega$ and let $x^0$ be a point with the property

$$\Omega \subset \left\{ y : \; |x^0 - y| \leqslant \operatorname{diam}\Omega \right\} .$$

Such a point we shall call a *center* of $\Omega$. Since smooth cut-off functions are pointwise multipliers, cf. Remark 33, we can associate with any $f \in B^s_q(L_p(\Omega))$ a tempered distribution $g \in B^s_q(L_p)$ such that $g|_\Omega = f$ in $\mathcal{D}'(\Omega)$,

$$c\|g\,|B^s_q(L_p)\| \leqslant \|f\,|B^s_q(L_p(\Omega))\| \leqslant \|\,g\,|B^s_q(L_p)\|, \tag{94}$$

$$\operatorname{supp} g \subset \{x \in \mathbb{R}^d : \quad |x - x^0| \leqslant 2\operatorname{diam}\Omega\}\,. \tag{95}$$

Here $0 < c < 1$ does not depend on $f$ (but on $\Omega, s, p, q$).

Now we turn to decompositions by means of wavelets. We use the notation from the preceeding subsection. Define

$$\Lambda_j := \left\{ k \in \mathbb{Z}^d : \quad |k_i - x_i^0| \leqslant 2^j \operatorname{diam} \Omega + N \,,\; i = 1, \ldots , d \right\} \,,\; j = 0, 1, \ldots . \tag{96}$$

Then given $f$ and taking $g$ as above, we find that

$$g = \sum_{k \in \Lambda_0} \langle g, \widetilde{\varphi}_k \rangle \, \varphi_k + \sum_{i=1}^{2^d-1} \sum_{j=0}^{\infty} \sum_{k \in \Lambda_j} \langle g, \widetilde{\psi}_{i,j,k} \rangle \, \psi_{i,j,k} \quad \text{(convergence in} \quad \mathcal{S}') \tag{97}$$

and

$$\begin{aligned} \|g \,|B_q^s(L_p)\| \asymp & \left( \sum_{k \in \Lambda_0} |\langle g, \widetilde{\varphi}_k \rangle|^p \right)^{1/p} \\ & + \left( \sum_{i=1}^{2^d-1} \sum_{j=0}^{\infty} 2^{jq(s+d(\frac{1}{2}-\frac{1}{p}))} \left( \sum_{k \in \Lambda_j} |\langle g, \widetilde{\psi}_{i,j,k} \rangle|^p \right)^{q/p} \right)^{1/q} < \infty \,. \end{aligned} \tag{98}$$

The following more handy notation is also used. We put

$$\nabla_{-1} := \Lambda_0, \tag{99}$$

$$\nabla_j := \left\{ (i,k) : \quad 1 \leqslant i \leqslant 2^d - 1 \,, \quad k \in \Lambda_j \right\} \,, \quad j = 0, 1, \ldots \,, \tag{100}$$

$\psi_{j,\lambda} := \psi_{i,j,k}$, if $\lambda = (i,k) \in \nabla_j$, $j \in \mathbb{N}_0$, and $\psi_{j,\lambda} := \varphi_k$ if $\lambda = k \in \nabla_{-1}$. For the dual basis, (97) and (98) read as

$$g = \sum_{j=-1}^{\infty} \sum_{\lambda \in \nabla_j} \langle g, \widetilde{\psi}_{j,\lambda} \rangle \, \psi_{j,\lambda} \quad \text{(convergence in } \mathcal{S}') \tag{101}$$

and

$$\|g \,|B_q^s(L_p)\| \asymp \left( \sum_{j=-1}^{\infty} 2^{jq(s+d(\frac{1}{2}-\frac{1}{p}))} \left( \sum_{\lambda \in \nabla_j} |\langle g, \widetilde{\psi}_{j,\lambda} \rangle|^p \right)^{q/p} \right)^{1/q} < \infty \,. \tag{102}$$

### *A.5. Lipschitz domains, embeddings, and interpolation*

We call a domain $\Omega$ a *special Lipschitz domain* (see [77]), if $\Omega$ is an open set in $\mathbb{R}^d$ and if there exists a function $\omega : \mathbb{R}^{d-1} \to \mathbb{R}$ such that

$$\Omega = \left\{ (x', x_d) \in \mathbb{R}^d \,:\, x_d > \omega(x') \right\}$$

and

$$|\omega(x') - \omega(y')| \leqslant c \, |x' - y'| \quad \text{for all } x',\ y' \in \mathbb{R}^{d-1} \,,$$

and some constant $c > 0$. We call a domain $\Omega$ a *bounded Lipschitz domain* if $\Omega$ is bounded and its boundary $\partial\Omega$ can be covered by a finite number of open balls $B_k$, so that, possibly after a proper rotation, $\partial\Omega \cap B_k$ for each $k$ is a part of the graph of a Lipschitz function.

**Proposition 6.** *Let $\Omega \in \mathbb{R}^d$ be a bounded Lipschitz domain with center $x^0$. Then there exists a universal bounded linear extension operator $\mathcal{E}$ for all values of $s$, $p$, and $q$, i.e.,*

$$(\mathcal{E} f)|_\Omega = f \qquad \textit{for all } f \in B^s_q(L_p(\Omega))$$

*and*

$$\| \mathcal{E} \, : \, B^s_q(L_p(\Omega)) \to B^s_q(L_p(\mathbb{R}^d)) \, \| < \infty.$$

*In addition we may assume*

$$\operatorname{supp} \mathcal{E} f \ \subset \{x \in \mathbb{R}^d : \quad |x - x^0| \leqslant 2 \operatorname{diam} \Omega\} \, . \tag{103}$$

**Remark 35.** *Proposition* 6 *has been proved by Rychkov* [75]. *Property* (103) *follows from Remark* 33.

Let us now discuss some embedding properties of Besov spaces that are needed for our purposes.

**Proposition 7.** *Let $\Omega \subset \mathbb{R}^d$ be an bounded open set. Let $0 < p_0, p_1, q_0, q_1 \leqslant \infty$ and let $s, t \in \mathbb{R}$. Then the embedding*

$$I \, : \, B^{s+t}_{q_0}(L_{p_0}(\Omega)) \to B^s_{q_1}(L_{p_1}(\Omega))$$

*is compact if and only if*

$$t > d \left( \frac{1}{p_0} - \frac{1}{p_1} \right)_+ \, . \tag{104}$$

**Remark 36.** Sufficiency is proved e.g., in [43]. The necessity of the given restrictions is almost obvious, but see Lemma 4 and [57] for details.

Sometimes Besov spaces or Sobolev spaces of fractional order are introduced by means of interpolation (real and/or complex). Here we state the following, cf. [91]. As usual, $(\,\cdot\,,\,\cdot\,)_{\Theta,q}$ and $[\,\cdot\,,\,\cdot\,]_\Theta$ denote the real and the complex interpolation functor, respectively.

**Proposition 8.** *Let $\Omega$ be a bounded Lipschitz domain. Let $0 < q_0, q_1 \leqslant \infty$ and let $s_0, s_1 \in \mathbb{R}$. Let $0 < \Theta < 1$.*

(i) *Let $0 < p, q \leqslant \infty$. Suppose $s_0 \neq s_1$ and put $s = (1 - \Theta)\, s_0 + \Theta s_1$. Then*

$$\Big( B^{s_0}_{q_0}(L_p(\Omega)), B^{s_1}_{q_1}(L_p(\Omega)) \Big)_{\Theta,q} = B^s_q(L_p(\Omega)) \quad (\textit{equivalent quasi-norms}).$$

(i) (ii) *Let $0 < p_0, p_1 \leqslant \infty$. We put $s = (1 - \Theta)\, s_0 + \Theta s_1$,*

$$\frac{1}{p} = \frac{1-\Theta}{p_0} + \frac{\Theta}{p_1} \quad \textit{and} \quad \frac{1}{q} = \frac{1-\Theta}{q_0} + \frac{\Theta}{q_1} \, .$$

*Then*

$$\Big[ B^{s_0}_{q_0}(L_{p_0}(\Omega)), B^{s_1}_{q_1}(L_{p_1}(\Omega)) \Big]_\Theta = B^s_q(L_p(\Omega)) \quad (\textit{equivalent quasi-norms}).$$

*A.6. Besov spaces on domains—intrinsic descriptions*

For $M \in \mathbb{N}$, $h \in \mathbb{R}^d$, and $f : \mathbb{R}^d \to \mathbb{C}$ we define

$$\Delta_h^M f(x) := \begin{cases} \sum_{j=0}^{M} \binom{M}{j}(-1)^{M-j} f(x+jh) & \text{if } x,\ x+h,\ \ldots,\ x+Mh \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

The corresponding modulus of smoothness is then given by

$$\omega^M(t,f)_p := \sup_{|h|<t} \| \Delta_h^M f \|_{L_p(\Omega)}, \qquad t > 0.$$

The approach by differences coincides with that using restrictions as can be seen by the recent result of Dispa [37].

**Proposition 9.** *Let $\Omega$ be a bounded Lipschitz domain. Let $M \in \mathbb{N}$. Let $0 < p, q \leqslant \infty$ and $d \max(0, 1/p - 1) < s < M$. Then*

$$B_q^s(L_p(\Omega)) = \left\{ f \in L_{\max(p,1)}(\Omega) : \right.$$
$$\left. \|f\|^{\Box} := \|f\|_{L_p(\Omega)} + \left( \int_0^1 \left[ t^{-s}\, \omega^M(t,f)_p \right]^q \frac{dt}{t} \right)^{1/q} < \infty \right\}$$

*in the sense of equivalent quasi-norms.*

*A.7. Sobolev spaces on domains*

Let $\Omega$ be a bounded Lipschitz domain. Let $m \in \mathbb{N}$. As usual $H^m(\Omega)$ denotes the collection of all functions $f$ such that the distributional derivatives $D^\alpha f$ of order $|\alpha| \leqslant m$ belong to $L_2(\Omega)$. The norm is defined as

$$\|f\,|H^m(\Omega)\| := \left( \sum_{|\alpha| \leqslant m} \|D^\alpha f\,|L_2(\Omega)\|^2 \right)^{1/2}.$$

It is well known that $H^m(\mathbb{R}^d) = B_2^m(L_2(\mathbb{R}^d))$ in the sense of equivalent norms, cf. e.g., [89]. As a consequence of the existence of a bounded linear extension operator for Sobolev spaces on bounded Lipschitz domains, cf. [77, p. 181], it follows that

$$H^m(\Omega) = B_2^m(L_2(\Omega)) \quad \text{(equivalent norms)}$$

for such domains. For fractional $s > 0$ we introduce the classes by complex interpolation. Let $0 < s < m$, $s \notin \mathbb{N}$. Then, following [59, 9.1], we define

$$H^s(\Omega) := \left[ H^m(\Omega), L_2(\Omega) \right]_\Theta, \quad \Theta = 1 - \frac{s}{m}.$$

This definition does not depend on $m$ in the sense of equivalent norms. This follows immediately from

$$\big[H^m(\Omega), L_2(\Omega)\big]_\Theta = \Big[B_2^m(L_2(\Omega)), B_2^0(L_2(\Omega))\Big]_\Theta = B_2^s(L_2(\Omega)), \qquad \Theta = 1 - \frac{s}{m}.$$

(all in the sense of equivalent norms), cf. Proposition 8.

### A.8. Function spaces on domains and boundary conditions

We concentrate on homogeneous boundary conditions. Here it makes sense to introduce two further scales of function spaces (distribution spaces).

**Definition 6.** Let $\Omega \subset \mathbb{R}^d$ be an open nontrivial set. Let $s \in \mathbb{R}$ and $0 < p, q \leqslant \infty$.

(i) Then $\mathring{B}_q^s(L_p(\Omega))$ denotes the closure of $\mathcal{D}(\Omega)$ in $B_q^s(L_p(\Omega))$, equipped with the quasi-norm of $B_q^s(L_p(\Omega))$.
(ii) Let $s \geqslant 0$. Then $H_0^s(\Omega)$ denotes the closure of $\mathcal{D}(\Omega)$ in $H^s(\Omega)$, equipped with the norm of $H^s(\Omega)$.
(iii) By $\widetilde{B}_q^s(L_p(\Omega))$ we denote the collection of all $f \in \mathcal{D}'(\Omega)$ such that there is a $g \in B_q^s(L_p(\mathbb{R}^d))$ with

$$g_{|\Omega} = f \quad \text{and} \quad \operatorname{supp} g \subset \overline{\Omega}, \tag{105}$$

equipped with the quasi-norm

$$\| f \,|\widetilde{B}_q^s(L_p(\Omega))\| = \inf \| g \,| B_q^s(L_p(\mathbb{R}^d))\|,$$

where the infimum is taken over all such distributions $g$ as in (105).

**Remark 37.** For a bounded Lipschitz domain $\mathring{B}_q^s(L_p(\Omega)) = \widetilde{B}_q^s(L_p(\Omega)) = B_q^s(L_p(\Omega))$ holds if

$$0 < p, q < \infty, \quad \max\left(\frac{1}{p} - 1, \, d\left(\frac{1}{p} - 1\right)\right) < s < \frac{1}{p},$$

cf. [48, Corollary 1.4.4.5, 91]. Hence,

$$H_0^s(\Omega) = \mathring{B}_2^s(L_2(\Omega)) = \widetilde{B}_2^s(L_2(\Omega)) = B_2^s(L_2(\Omega)) = H^s(\Omega)$$

if $0 \leqslant s < 1/2$.

Often it is more convenient to work with a scale $\overline{B}_q^s(L_p(\Omega))$, originally introduced in [91].

**Definition 7.** Let $\Omega \subset \mathbb{R}^d$ be an open nontrivial set. Let $s \in \mathbb{R}$ and $0 < p, q \leqslant \infty$. Then we put

$$\overline{B}_q^s(L_p(\Omega)) := \begin{cases} B_q^s(L_p(\Omega) & \text{if } s < 1/p, \\ \widetilde{B}_q^s(L_p(\Omega)) & \text{if } s \geqslant 1/p. \end{cases}$$

This scale $\overline{B}_q^s(L_p(\Omega))$ is well behaved under interpolation and duality, cf. [91].

**Proposition 10.** *Let $\Omega$ be a bounded Lipschitz domain. Let $1 < p, p_0, p_1, q, q_0, q_1 < \infty$ and let $s_0, s_1 \in \mathbb{R}$. Let $0 < \Theta < 1$.*

(i) *Suppose $s_0 \neq s_1$ and put $s = (1-\Theta)\, s_0 + \Theta\, s_1$. Then*

$$\left(\overline{B}^{s_0}_{q_0}(L_p(\Omega)), \overline{B}^{s_1}_{q_1}(L_p(\Omega))\right)_{\Theta,q} = \overline{B}^{s}_{q}(L_p(\Omega)) \quad (\textit{equivalent quasi-norms}).$$

(ii) *We put $s = (1-\Theta)s_0 + \Theta s_1$,*

$$\frac{1}{p} = \frac{1-\Theta}{p_0} + \frac{\Theta}{p_1} \qquad \textit{and} \qquad \frac{1}{q} = \frac{1-\Theta}{q_0} + \frac{\Theta}{q_1}\,.$$

*Then*

$$\left[\overline{B}^{s_0}_{q_0}(L_{p_0}(\Omega)), \overline{B}^{s_1}_{q_1}(L_{p_1}(\Omega))\right]_{\Theta} = \overline{B}^{s}_{q}(L_p(\Omega)) \quad (\textit{equivalent quasi-norms}).$$

(iii) *With $s \in \mathbb{R}$ and*

$$1 = \frac{1}{p} + \frac{1}{p'} \quad \textit{and} \quad 1 = \frac{1}{q} + \frac{1}{q'}$$

*we find*

$$\left(\overline{B}^{s}_{q}(L_p(\Omega))\right)' = \overline{B}^{-s}_{q'}(L_{p'}(\Omega))\,.$$

*Here the duality must be understood in the framework of the dual pairing $(\mathcal{D}(\Omega), \mathcal{D}'(\Omega))$.*

*A.9. Sobolev spaces with negative smoothness*

**Definition 8.** For $s > 0$ we define

$$H^{-s}(\Omega) := \begin{cases} \left(H^s_0(\Omega)\right)' & \text{if } s - \frac{1}{2} \neq \text{integer}, \\ \left(\widetilde{B}^s_2(L_2(\Omega))\right)' & \text{otherwise.} \end{cases}$$

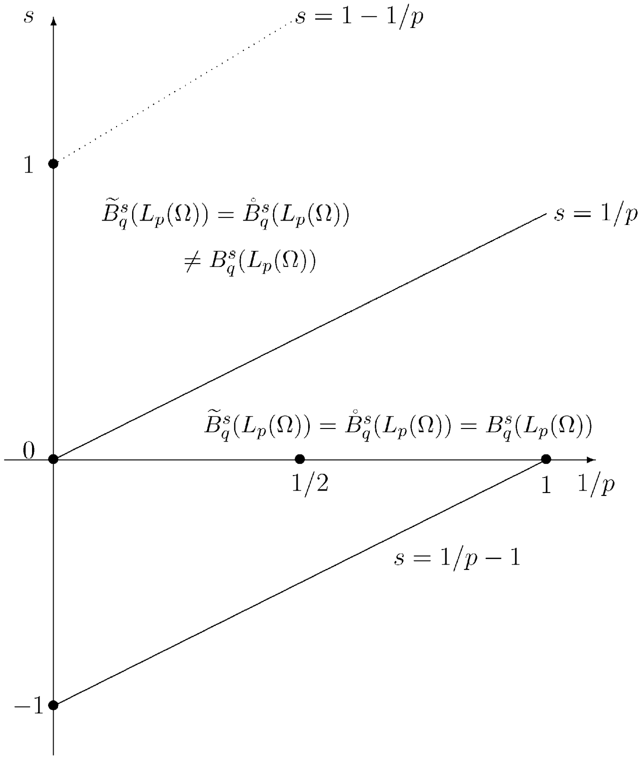
**Remark 38.** Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Then

$$H^s_0(\Omega) = \widetilde{B}^s_2(L_2(\Omega)), \quad s > 0, \;\; s - \tfrac{1}{2} \neq \text{integer}\,,$$

cf. [48, Corollary 1.4.4.5] and Proposition 9. From Remark 37 and Proposition 10 we conclude the identity

$$H^{-s}(\Omega) = B^{-s}_2(L_2(\Omega)), \quad s > 0, \tag{106}$$

to be understood in the sense of equivalent norms.




**Remark 39** (*Triebel [88, 4.3.2]*). Let $\Omega$ be a bounded open set with a smooth boundary. Then $\mathring{B}^s_q(L_p(\Omega)) = \widetilde{B}^s_q(L_p(\Omega))$ holds if

$$1 < p, q < \infty, \quad \frac{1}{p} - 1 < s < \infty, \quad s - \frac{1}{p} \neq \text{integer}.$$

*A.10. Wavelet characterization of Besov spaces on domains*

It is a difficult task to construct wavelet bases on domains, see [12, 2.12] and the references given there. Under certain conditions on the domain $\Omega$ such constructions with properties similar to (101), (102) are known in the literature, see Remark 11.

Let $\Omega$ be a bounded open set in $\mathbb{R}^d$. Let $p$, $q$ and $s$ be fixed such that $s > d \max(0, 1/p - 1)$. We suppose that there exist sets $\nabla_j \subset \{1, 2, \dots, 2^d - 1\} \times \mathbb{Z}^d$, with

$$0 < \inf_{j=-1,0,\dots} 2^{-jd} \, |\nabla_j| \leqslant \sup_{j=-1,0,\dots} 2^{-jd} \, |\nabla_j| < \infty, \tag{107}$$

and functions $\psi_{j,\lambda}$, $\widetilde{\psi}_{j,\lambda}$, $\lambda \in \nabla_j$, $j = -1, 0, 1, \dots$, such that

$$\operatorname{supp} \psi_{j,\lambda}, \quad \operatorname{supp} \widetilde{\psi}_{j,\lambda} \subset \Omega, \quad \lambda \in \nabla_j, \tag{108}$$

$$\langle \widetilde{\psi}_{i,j,k}, \psi_{u,v,\ell} \rangle = \delta_{i,u} \, \delta_{j,v} \, \delta_{k,\ell}, \tag{109}$$

and such that $f \in B^s_q(L_p(\Omega))$ if and only if

$$f = \sum_{j=-1}^{\infty} \sum_{\lambda \in \nabla_j} \langle f, \widetilde{\psi}_{j,\lambda} \rangle \, \psi_{j,\lambda} \quad \text{(convergence in } \mathcal{D}'\text{)}\,, \tag{110}$$

and

$$\|f\|^{\clubsuit}_{B^s_q(L_p(\Omega))} \asymp \|f\|_{B^s_q(L_p(\Omega))}, \tag{111}$$

where

$$\|f\|^{\clubsuit}_{B^s_q(L_p(\Omega))} := \left( \sum_{j=-1}^{\infty} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))q} \left( \sum_{\lambda \in \nabla_j} |\langle f, \widetilde{\psi}_{j,\lambda} \rangle|^p \right)^{q/p} \right)^{1/q} < \infty\,. \tag{112}$$

## Acknowledgments

We thank Stefan Heinrich, Peter Mathé, Volodya Temlyakov, Hans Triebel, and Art Werschulz for many valuable remarks and comments.

## References

[1] I. Babuška, W.C. Rheinboldt, A posteriori error estimates for finite element methods, Internat. J. Numer. Math. Eng. 12 (1978) 1597–1615.
[2] N.S. Bakhvalov, On the optimality of linear methods for operator approximation in convex classes of functions, USSR Comput. Math. and Math. Phys. 11 (1971) 244–249.
[3] R.E. Bank, A. Weiser, Some a posteriori error estimators for elliptic partial differential equations, Math. Comp. 44 (1985) 283–301.
[4] R. Becker, C. Johnson, R. Rannacher, Adaptive error control for multi-grid finite element methods, Computing 55 (1995) 271–288.
[5] F. Bornemann, B. Erdmann, R. Kornhuber, A posteriori error estimates for elliptic problems in two and three space dimensions, SIAM J. Numer. Anal. 33 (1996) 1188–1204.
[6] G. Bourdaud, Ondelletes et des espaces de Besov, Rev. Mat. Iberoamericana 11 (1995) 477–512.
[7] A.M. Caetano, About approximation numbers in function spaces, J. Approx. Theory 94 (1998) 383–395.
[8] C. Canuto, A. Tabacco, K. Urban, The wavelet element method, part I: construction and analysis, Appl. Comput. Harmonic Anal. 6 (1999) 1–52.
[9] B. Carl, Entropy numbers, $s$-numbers and eigenvalue problems, J. Funct. Anal. 41 (1981) 290–306.
[10] B. Carl, I. Stephani, Entropy, Compactness and the Approximation of Operators, Cambridge University Press, Cambridge, 1990.
[11] Z. Ciesielski, T. Figiel, Spline bases in classical function spaces on compact $C^{\infty}$ manifolds, part 1 and 2, Studia Math. 76 (1983) 1–58, 95–36.
[12] A. Cohen, Numerical Analysis of Wavelet Methods, Elsevier, Amsterdam, 2003.
[13] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods for elliptic operator equations—convergence rates, Math. Comp. 70 (2001) 22–75.
[14] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet methods II—beyond the elliptic case, Found. Comput. Math. 2 (2002) 203–245.
[15] A. Cohen, W. Dahmen, R. DeVore, Adaptive wavelet schemes for nonlinear variational problems, SIAM J. Numer. Anal. 41 (5) (2003) 1785–1823.
[16] S. Dahlke, Besov regularity for elliptic boundary value problems with variable coefficients, Manuscripta Math. 95 (1998) 59–77.
[17] S. Dahlke, Besov regularity for elliptic boundary value problems in polygonal domains, Appl. Math. Lett. 12 (6) (1999) 31–38.

[18] S. Dahlke, Besov regularity for interface problems, Z. Angew. Math. Mech. 79 (1999) 383–388.
[19] S. Dahlke, Besov regularity for the Stokes problem, in: W. Haussmann, K. Jetter, M. Reimer (Eds.), Advances in Multivariate Approximation, Mathematical Research, vol. 107, Wiley-VCH, Berlin, 1999, pp. 129–138.
[20] S. Dahlke, W. Dahmen, R. DeVore, Nonlinear approximation and adaptive techniques for solving elliptic operator equations, in: W. Dahmen, A. Kurdila, P. Oswald (Eds.), Multiscale Wavelet Methods for Partial Differential Equations, Academic Press, San Diego, 1997, pp. 237–283.
[21] S. Dahlke, W. Dahmen, R. Hochmuth, R. Schneider, Stable multiscale bases and local error estimation for elliptic problems, Appl. Numer. Math. 23 (1997) 21–48.
[22] S. Dahlke, W. Dahmen, K. Urban, Adaptive wavelet methods for saddle point problems—optimal convergence rates, SIAM J. Numer. Anal. 40 (4) (2002) 1230–1262.
[23] S. Dahlke, R. DeVore, Besov regularity for elliptic boundary value problems, Comm. Partial Differential Equations 22 (1&2) (1997) 1–16.
[24] S. Dahlke, E. Novak, W. Sickel, Optimal approximation of elliptic problems by linear and nonlinear mappings I, J. Complexity 22 (2006) 29–49.
[26] W. Dahmen, R. Schneider, Composite wavelet bases for operator equations, Math. Comp. 68 (1999) 1533–1567.
[27] W. Dahmen, R. Schneider, Wavelets on manifolds I: construction and domain decomposition, SIAM J. Math. Anal. 31 (1999) 184–230.
[28] K. Deimling, Nonlinear Functional Analysis, Springer, Berlin, 1985.
[29] R.A. DeVore, Nonlinear approximation, Acta Numerica 7 (1998) 51–150.
[30] R.A. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, Manuscripta Math. 63 (1989) 469–478.
[31] R.A. DeVore, B. Jawerth, V. Popov, Compression of wavelet decompositions, Amer. J. Math. 114 (1992) 737–785.
[32] R.A. DeVore, G. Kyriazis, D. Leviatan, V.M. Tikhomirov, Wavelet compression and nonlinear $n$-widths, Adv. Comput. Math. 1 (1993) 197–214.
[33] R.A. DeVore, G. Petrova, V. Temlyakov, Best basis selection for approximation in $L_p$, Found. Comput. Math. 3 (2003) 161–185.
[34] R.A. DeVore, V. Popov, Interpolation spaces and nonlinear approximation, in: Function Spaces and Approximation, Lecture Notes in Mathematics, vol. 1302, 1988, pp. 191–205.
[35] R.A. DeVore, R.C. Sharpley, Besov spaces on domains in $\mathbb{R}^d$, Trans. Amer. Math. Soc. 335 (1993) 843–864.
[36] R.A. DeVore, V.N. Temlyakov, Nonlinear approximation by trigonometric sums, J. Fourier Anal. Appl. 2 (1995) 29–48.
[37] S. Dispa, Intrinsic characterizations of Besov spaces on Lipschitz domains, Math. Nachr. 260 (2002) 21–33.
[38] D. Donoho, Unconditional bases are optimal for data compression and for statistical estimation, Appl. Comput. Harmonic Anal. 1 (1993) 100–115.
[39] W. Dörfler, A convergent adaptive algorithm for Poisson's equation, SIAM J. Numer. Anal. 33 (1996) 737–785.
[40] D. Dung, V.Q. Thanh, On nonlinear $n$-widths, Proc. Amer. Math. Soc. 124 (1996) 2757–2765.
[41] D. Dung, Continuous algorithms in $n$-term approximation and non-linear widths, J. Approx. Theory 102 (2000) 217–242.
[43] D.E. Edmunds, H. Triebel, Function Spaces, Entropy Numbers, Differential Operators, Cambridge University Press, Cambridge, 1996.
[44] S. Gal, C.A. Micchelli, Optimal sequential and non-sequential procedures for evaluating a functional, Appl. Anal. 10 (1980) 105–120.
[45] E.D. Gluskin, On some finite dimensional problems of the theory of diameters, Vestnik Leningr. Univ. 13 (1981) 5–10.
[46] E.D. Gluskin, Norms of random matrices and diameters of finite dimensional sets, Mat. Sb. 120 (1983) 180–189.
[48] P. Grisvard, Elliptic Problems in Nonsmooth Domains, Pitman, Boston, 1985.
[49] P. Grisvard, Singularities in Boundary Value Problems, Research Notes in Applied Mathematics, vol. 22, Springer, Berlin, 1992.
[50] W. Hackbusch, Elliptic Differential Equations: Theory and Numerical Treatment, Springer, Berlin, 1992.
[51] S. Heinrich, The randomized information complexity of elliptic PDE, J. Complexity 22 (2006) 220–249.
[52] S. Heinrich, The quantum query complexity of elliptic PDE, Preprint, 2006.
[54] B. Kashin, Approximation properties of complete orthonormal systems, Trudy Mat. Inst. Steklov 172 (1985) 187–191.
[56] T. Kuhn, H.-G. Leopold, W. Sickel, L. Skrzypczak, Entropy numbers of embeddings of weighted Besov spaces. II, Proc. Edinburgh Math. Soc. (2003), to appear.
[57] H.-G. Leopold, Embeddings for general weighted sequence spaces and entropy numbers, in: Function Spaces, Differential Operators, and Nonlinear Analysis, Academy of Sciences of the Czech Republic, Praha 2000, 1999, pp. 170–186.

[58] R. Linde, $s$-Numbers of diagonal operators and Besov embeddings, in: Proceedings of the 13th Winter School on Abstract Analysis, Rend. Circ. Mat. Palermo Ser. II. Suppl. 10 (1985) 83–110.
[59] J.L. Lions, E. Magenes, Non-Homogeneous Boundary Value Problems and Applications I, Springer, Berlin, 1972.
[60] P. Mathé, $s$-Numbers in information-based complexity, J. Complexity 6 (1990) 41–66.
[61] J.M. Melenk, On $n$-widths for elliptic problems, J. Math. Anal. Appl. 247 (2000) 272–289.
[62] Y. Meyer, Wavelets and Operators, Cambridge University Press, Cambridge, 1992.
[63] S.M. Nikol'skij, Approximation of Functions of Several Variables and Imbedding Theorems, Springer, Berlin, 1975.
[64] E. Novak, The real number model in numerical analysis, J. Complexity 11 (1995) 57–73.
[65] E. Novak, On the power of adaption, J. Complexity 12 (1996) 199–237.
[66] E. Novak, H. Woźniakowski, On the cost of uniform and nonuniform algorithms, Theoret. Comput. Sci. 219 (1999) 301–318.
[67] E. Novak, H. Woźniakowski, Complexity of linear problems with a fixed output basis, J. Complexity 16 (2000) 333–362.
[68] P. Oswald, On the degree of nonlinear spline approximation in Besov–Sobolev spaces, J. Approx. Theory 61 (1990) 131–157.
[69] J. Peetre, New Thoughts on Besov Spaces, Duke Univ. Math. Ser., Durham, 1976.
[71] A. Pietsch, $s$-Numbers of operators in Banach spaces, Studia Math. 51 (1974) 201–223.
[72] A. Pietsch, Eigenvalues and $s$-Numbers, Geest und Portig, Leipzig, 1987.
[73] A. Pinkus, $n$-Widths in Approximation Theory, Springer, Berlin, 1985.
[74] T. Runst, W. Sickel, Sobolev Spaces of Fractional Order, Nemytskij Operators and Nonlinear Partial Differential Equations, de Gruyter, Berlin, 1996.
[75] V.S. Rychkov, On restrictions and extensions of the Besov and Triebel–Lizorkin spaces with respect to Lipschitz domains, J. London Math. Soc. 60 (1999) 237–257.
[77] E.M. Stein, Singular Integrals and Differentiability Properties of Functions, Princeton University Press, Princeton, NJ, 1989.
[78] A.I. Stepanets, Approximation characteristics of the spaces $S_{\varphi}^{p}$ in different metrics, Ukrainian Math. J. 53 (2001) 1340–1374.
[79] M.I. Stesin, Aleksandrov diameters of finite-dimensional sets and of classes of smooth functions, Dokl. Akad. Nauk SSSR 220 (1974) 1278–1281.
[80] R. Stevenson, Optimality of a standard adaptive finite element method, Preprint 1329, Department of Mathematics, Utrecht University, 2005.
[81] V.N. Temlyakov, Approximation of Periodic Functions, Nova Science, New York, 1993.
[82] V.N. Temlyakov, Greedy algorithms with regard to multivariate systems with special structure, Constr. Approx. 16 (2000) 399–425.
[83] V.N. Temlyakov, Universal bases and greedy algorithms for anisotropic function classes, Constr. Approx. 18 (2002) 529–550.
[84] V.N. Temlyakov, Nonlinear methods of approximation, Found. Comput. Math. 3 (2003) 33–107.
[85] V.M. Tikhomirov, Approximation Theory, Encyclopaedia of Mathematical Sciences Analysis II, vol.14, Springer, Berlin, 1990.
[86] J.F. Traub, G.W. Wasilkowski, H. Wozniakowski, Information-Based Complexity, Academic Press, New York, 1980.
[87] J.F. Traub, H. Woźniakowski, A General Theory of Optimal Algorithms, Academic Press, New York, 1980.
[88] H. Triebel, Interpolation Theory, Function Spaces, Differential Operators, VEB Deutscher Verlag der Wissenschaften, Berlin, 1978.
[89] H. Triebel, Theory of Function Spaces, Birkhäuser, Basel, 1983.
[90] H. Triebel, Theory of Function Spaces. II, Birkhäuser, Basel, 1992.
[91] H. Triebel, Function spaces in Lipschitz domains and on Lipschitz manifolds. Characteristic functions as pointwise multipliers, Rev. Mat. Complutense 15 (2002) 475–524.
[92] H. Triebel, A note on wavelet bases in function spaces. Proceedings of the Orlicz Centenary Conference on Function Spaces, vol. 7, Banach Center Publications, vol. 64, Polish Academy of Sciences, 2004, pp. 193–206.
[93] R. Verfürth, A posteriori error estimation and adaptive mesh-refinement techniques, J. Comput. Appl. Math. 50 (1994) 67–83.
[94] A.G. Werschulz, The Computational Complexity of Differential and Integral Equations, Oxford Science Publications, Oxford, 1996.